The way researchers work with data is shifting dramatically, driven in large part by the growing popularity of the programming language R. Initially released in 1996, R has become the analytical tool of choice for many in research, academia and business because:

1. R is a free, open-source package available for Linux, Windows and Mac.
2. R is supported by an extremely active community of users who develop new tools and share their work as “libraries”. As a result, R is a dynamic, growing resource, providing access to the most current methods for data manipulation, analysis and presentation.
3. Unlike commercial packages, open source software is highly adaptable. A routine written by one user can easily be adopted, or adapted for use by others.
4. R is managed by a leadership team that oversees code development, releasing updates on a 6-month cycle. New methods of data management, analysis and presentation are thus available as they are developed (in contrast to the years it may take for developments to appear in commercial packages). A web search reveals several thousand freely distributed R libraries related to data management, univariate and multivariate statistical analysis, evaluation of complex genomic sequences, creation of publication-quality graphics, etc.

R’s adoption by companies like Google, LinkedIn, Facebook, Bank of America and Shell indicate that the program is here to stay, and will become increasingly influential in the world of data analysis (Vance 2009, Muenchen n.d.). A 2012 survey done by the data mining site KDnuggets (KDnuggets n.d.) suggests that R is becoming the primary choice of many data analysts.

Given the increasing importance of R, it is clear that university students will benefit enormously from familiarity with the program, particularly if taught in the context of topics they will continue to use after graduation. I propose to fundamentally change ESCI 502 (Experimental Design), to incorporate R as the primary analytical/teaching tool.

ESCI 502, generally taken during a graduate student’s first quarter at WWU, is core curriculum for many Environmental Science and Biology graduate students (and a handful of motivated undergraduates). The short-term course goal is to teach concepts of experimental design and data analysis before students start their graduate research. The longer-term goal is to produce well-trained data consumers/analysts prepared to enter the workforce and contribute to the field. ESCI 502 course evaluations are consistently strong (Table 1). Written comments indicate that students value the high-level treatment of statistical principles and the practical, hands-on exposure to software and analytical techniques they receive.

Table 1. Selected results from ESCI 502 evaluations (2003-2012). Averages are shown with 5.0 the highest possible value.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intellectual challenge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>offered by the course was</td>
<td>4.86</td>
<td>4.76</td>
<td>4.91</td>
<td>4.89</td>
<td>4.89</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>4.94</td>
<td>4.86</td>
</tr>
<tr>
<td>The overall course was</td>
<td>4.63</td>
<td>4.48</td>
<td>4.64</td>
<td>4.89</td>
<td>4.89</td>
<td>4.89</td>
<td>4.90</td>
<td>4.71</td>
<td>4.88</td>
<td>4.92</td>
</tr>
</tbody>
</table>
A critical aspect of ESCI 502 is giving students practical experience using software for data manipulation, analysis and display. I accomplish this through demonstration, classroom discussion, lab tutorials, weekly homework assignments and a 235-page course manual that includes presentation of complete data sets, analyses and annotated output for > 25 statistical methods ranging from simple ANOVA to multi-dimensional scaling. Feedback from course alumni indicates that this manual is a critical resource years after the course is over. The foundation of the class has been the menu-driven statistical package SPSS (licensed for campus labs but too expensive for student home use), supplemented by PRIMER-E ($500/license) and a very old DOS program for analyses not available in either of those packages. SigmaPlot ($549/license) serves for graphics in some analyses. Because most of this software is expensive and available on a limited number of university computers, I struggle to provide adequate access to the students. Furthermore, once the students leave WWU, they usually have no further access to any of these packages. It is clearly time for a fundamental change. R, which the students can freely download to their personal, laboratory or workplace computers, can replace all of this software, giving students unlimited access to the tools we use in class and providing continuity once they leave the university. ESCI 502 will take a significant leap forward in relevance if the entire foundation can be shifted to R. This will require 1) increasing my own familiarity with the program to the point I can seamlessly incorporate it into my teaching, 2) developing R approaches for each statistical principle I teach, 3) rewriting each lecture/discussion, 4) redoing the homework assignments, and 5) rewriting the ESCI course manual.

Because ESCI 502 treats such a range of statistical methods, transitioning from SPSS to R will take a great deal of work, particularly since the methods we use in class go well beyond the simple introductory statistics generally taught. I am still learning the program myself, but there are a number of excellent R reference books (e.g., Muenchen 2011, Crawley 2013) and some very active R forums and discussion groups. I also propose to take an online CourseRA statistics course (by Andrew Conway at Princeton) that promises to provide additional good R exposure. Some initial work suggests that every method I teach can be done with R, but that it will take some time to find the approaches most accessible to the students. I piloted several R exercises this summer while teaching an undergraduate statistics course for the University of Oregon. The results were very encouraging.

Incorporating R into ESCI 502 and, by extension, into my other courses (e.g. ESCI 417: Research in Environmental Science, ESCI 445: Current Trends in Marine Science) will permit me to teach students a tool that is eclipsing traditional software programs. The general availability of R will improve learning as all students will have similar access for homework and study. They will come away better prepared to work with their own data and more competitive for the ever increasing number of professional positions that require R experience. Having time to focus on R is also key to my own professional development and my ability to keep pace with a rapidly developing field. I look forward to this opportunity to upgrade my skills as an instructor, researcher and mentor to undergraduate and graduate students preparing for careers in biology and environmental science.
Literature Cited


