POLITICAL SCIENCE 900 SYLLABUS: APPLIED STATISTICAL PROGRAMMING FOR SOCIAL SCIENTISTS

Fall 2017

Instructor:	Shahryar Minhas	**Time:**	R 12:40pm – 3:30pm
Teaching Assistant:	Zuhaib Mahmood		
Email:	minhassh@msu.edu	**Place:**	104 S. Kedzie
TA Email:	mahmoo21@msu.edu		
Office Hours:	T 9:00am – 11:00am	**Office:**	317 South Kedzie Hall
TA Office Hours:	TBD	**TA Office:**	TBD

Additional information and materials in relation to the class will be posted on D2L.

Course Description:

Statistical computing is a quickly changing field. Standard techniques of today would have been difficult to execute fifteen years ago and impossible even in the late 1990s. Rapid improvements in computing power have been accompanied by swift changes in standard statistical methods. In just the last decade, techniques ranging from MCMC simulation, spatial, network, and text analyses have moved from being novel, advanced applications to commonplace across the social sciences.

This class is designed to achieve two broad objectives. First, this course aims to give you the more foundational meta-skills from computer science and statistics you need to teach yourself how to develop code or even software to execute specific tasks in R or similar computer languages. In addition, the course will introduce you about how these foundational skills can be used to study complex data structures that are becoming increasingly common in political science.

Grading

Your final grade will be comprised of the following:

Mini Projects	40%	Final Project	40%
In-Class Midterm	10%	Class Engagement	10%

The grading scale will be as follows:

- 4.0 (92-100 percent)
- 3.5 (87-91.9 percent)
- 3.0 (80-86.9 percent)
- 2.5 (76-79.9 percent)
- 2.0 (70-75.9 percent)
- 1.5 (65-69.9 percent)
- 1.0 (59.5-64.9 percent)
- 0.0 (<59.5)

Teams

Each of you will be assigned to a team of ≈3 students. You and your team will work together on both in-class assignments and the mini projects. I will construct the teams based on your previous exposure to programming topics. Unless there is some issue, the teams will not change throughout the semester.

Lectures

Each class will involve me giving a lecture or walkthrough of how to handle some tasks in R, but they are also designed to be interactive. Throughout each class I will ask you to apply the tools and techniques that I have introduced with your team. Learning to work in a statistical computing language like R involves doing more than listening. As such this will also require me to enforce a fairly rigorous class attendance policy. If you are unable to attend a class, you must let me know ahead of time and have a valid reason. Please note that class engagement comprises ten percent of your final grade.

Mini Projects

Forty percent of your grade will be determined by your performance on a series of mini projects that you will complete in collaboration with your team. Since you are operating as a team everyone should contribute equally to each assignment, and at the end of each assignment you will be asked to evaluate your team members. Mini projects will be assigned roughly every other week and are expected to be completed by the due date. Late assignments will be penalized as follows:

- Late, but same day: -10%
- Late, next day: -20%
- 2 days or later: no credit

Midterm

There will be one in-class midterm on March 1st that you are expected to complete individually and will comprise ten percent of your grade. The exam will ask you to complete a number of programming tasks related to the material presented in the class by that date.

Final Project

Another forty percent of your grade will be determined by your final project. This will be a somewhat open-ended exercise BUT the end product of it must be a 10-12 double-spaced research paper that involves a significant amount of coding in R. You will be allowed to form your own team of up to 3 students for this paper. Use this final project to your advantage in terms of your prelims, dissertation, or a manuscript you're planning on submitting with colleagues. You must submit a proposal describing what you're doing by February 22. Following the delivery of the proposal, I will meet with you to discuss whether the proposal is satisfactory. Not turning the proposal in on time will lead to a -10% reduction in the grade for your final project. Final project and all code must be submitted by May 1st at 2:45pm on D2L.

Presentation of the Final Project

[thanks Connie!]

Each student will create a poster summarizing their research paper. A poster session will be held jointly with Connie's multilevel modeling class on April 27, 2018 (tentative). Faculty and students will be invited to attend. Presenters should stay near their posters to take questions and comments and explain their findings to attendees.

A draft of the poster should be presented to the instructor on (or before) April 12, 2018. On April 19, 2018, there will be a workshop during class in which students can work on their posters and papers: they can receive assistance from the instructor or other classmates, discuss questions or problems with their projects, seek suggestions on how to improve their work, etc. To take full advantage of the opportunity students need to have a complete draft of their project.

Take into consideration that posters should go into print at least 2 or 3 days before the poster session. The poster can be landscape or portrait, but no larger than 36 x 48 inches. The following provides helpful advice about structuring and organizing a good poster: Designing Effective Posters, Jeff Radel at the University of Kansas. There are a variety of software packages that can be used to design posters including Microsoft Power Point, TeX, and Adobe Illustrator.

Helpful Resources

No books. Here are some helpful resources:

- Wickham's Advanced R Book
- R Cookbook
- ggplot2
- stackoverflow

Class Equipment

You will need a computer that has some specific software on it. You should be able to do everything we do in class using modern Mac :), Linux :) and even Windows :(operating systems. If you have questions, ask me. In the course, we will be using free, open-source software. It will be necessary to have a computer with multiple cores and $\approx$ 128gb hard drive in class.

- R: This is a free program, available here.
- Some people find RStudio a very useful to interact with R.
- Github: Set up an account here github.

Policies

Missing class or midterm: If you know that you are going to have to miss a class or the midterm, let me know at least 48 hours beforehand. If you simply skip the midterm, you will get a zero and no make-up will be offered.

Academic Integrity

Don't cheat . . . `https://ombud.msu.edu/academic-integrity/student-faq.html`:

> "Academic honesty and integrity are fundamental values in a community of scholars. As stated in the MSU Student Rights and Responsibilities and Spartan Code of Honor, students and faculty share a commitment to and responsibility for "maintaining the integrity of scholarship, grades, and professional standards." To abuse these values is to assault one's own personal integrity and character.

Don't take without attribution . . . `http://www2.stat.duke.edu/~cr173/Sta323_Sp17/syllabus/`:

> "A note on sharing / reusing code – I am well aware that a huge volume of code is available on the web to solve any number of problems. Unless I explicitly tell you not to use something the course's policy is that you may make use of any online resources (e.g. StackOverflow) but you must explicitly cite where you obtained any code you directly use (or use as inspiration). Any recycled code that is discovered and is not explicitly cited will be treated as plagiarism. The one exception to this rule is that you may not directly share code with another team in this class, you are welcome to discuss the problems together and ask for advice, but you may not send or make use of code from another team."

Fixed Dates

Proposal for Final Project due on 2/22

Midterm on 3/1

No Class: Spring Break on 3/8

No Class: Traveling for a conference on 4/8

Poster Draft Due on 4/12

In-Class Poster Day on 4/19

Present Posters on 4/27

Final Project Due on 5/1 by 2:45pm

Approximate Class Schedule

Hi R!

- Introduction to the Syllabus
- R Markdown
- Github
- R
  - Arithmetic, 1-d objects, object manipulation

Objects, Objects, & Doing Stuff to Objects

- Adding dimensions and layers to objects
- Logical statements
- Loops

Embrace the Function-ality of R

- Functions
- Functional programming
- Test-driven development

Lurking Beneath the Surface

- Classes
- Environments

Pretty Pictures

- Plotting Principles
- Base graphics
- **ggplot2**

Inundated with Data ... Kinda

- Regular expressions
- Scraping Data from APIs like Twitter & Facebook: Sites like Twitter and Facebook offer a set of structured *http* requests that return JSON or XML files, this week we will learn to work with APIs such as these and will also discuss how social scientists have used data from these sites.
- Scraping Data from the Web: An increasing amount of data is available on the web (e.g., election results, budget allocations, legislative speeches). These data are provided in an unstructured format: you can always copy & paste, but it's time-consuming and prone to errors. Web scraping is the process of extracting this information automatically and transform it into a structured dataset.

Use All the Cores! & Forays into Geography

- Benchmarking
- Parallelization
- Spatial Data: Processes such as democratization and conflict have been argued to diffuse spatially. We will review how to visualize and extract descriptive information from spatial data using R.

Use All the Cores BETTER! & Forays into Networks

- Lower level computation using **Rcpp**
- Networks Application: Network analysis is becoming an increasingly common mode of analysis in the social sciences. We will review how network analysis has been used in the social sciences and how to work with networks in R.

Interactive Visualizations + Text ... there's a Connection

- *RShiny*: Interactive visualizations are a great way to better understand and explore your data. The team at RStudio have developed a package, **Shiny**, through which we can build interactive web applications straight from R.
- *Textual data*: Words matter in political science and in recent years political scientists have increasingly utilized data analytic techniques to study treaties, congressional bills, and legislative speeches. We will go over how to structure, visualize, and summarise textual data in R.

Organizing Academic Projects

- *Importance of reproducible research*: For your data analysis to be trusted, others must be able to reproduce the work that you have done.

The Cloud

- Cloud computing: Amazon Web Services, SparklyR

Share Your Brilliance

- How to build an R Package.