POLSCI 514 Winter - Intermediate Programming Topics for Social Research

Instructor: Fabricio Vasselai (Winter 2019)

Friday 10am-11am at room 7603, Haven Hall

E-mail: vasselai@umich.edu
Office: room 7740 at Haven Hall Office Hours: 3pm-5pm, Thursdays
Class at Canvas: https://umich.instructure.com/courses/271133
Class at Piazza: https://piazza.com/configure-classes/winter2019/polsci514

Content summary:

This class is a continuation to Polsci 514 Fall, where students were introduced to the basics of programming and to the basics of R syntax with a focus on loops, code control-flow, logical operators and implicit loops. This time, students will be introduced to different topics of real life research. While the entirety of the class will be taught using R, students that want to use this opportunity to learn either Python or C++ are allowed to submit assignments done with those programming languages as well[1]. If desired, I can point students towards good general material for learning Python and C++. Come talk to me if you consider focusing (at least part of the) the semester on one of those languages.

If you took Polsci 514 Fall, notice that the format of Polsci 514 Winter will be rather different. We will focus much less on R syntax and much more on real life usage, which requires this class to be organized on specific (often self-contained) topics. Accordingly, there will be (short) required readings almost every week, to guarantee that students come to class with necessary minimum background to understand the concepts behind what will be done with R. Because lectures happen only once a week and last only 1 hour, it is **crucial** that you read the required readings before attending lecture.

Prerequisites:

Because of the short frequency and duration of lectures, **there will not be time** for much review. Hence, since the beginning of the semester, it will be **assumed** that students have:

- R programming knowledge at the level of Polsci 514 Fall.
- Math at the level of Polsci 598.
- Statistics at the level of Polsci 599.

For the final project, it will also be assumed that students can typeset text in LaTeX.

Notice that while students are not required to be currently enrolled in Polsci 681 and Polsci 699, a few topics here relate to what is being taught there. Still, I will always provide readings to give enough context to students that are not taking those classes. Also notice that a couple of our lecture topics (see below) might change as a function of those classes, since I want the majority of the students to have been exposed to some theory before illustrating techniques (e.g. in the case of OLS regression).

[1] Notice that the challenge might be bigger with Python than with R and it will certainly be super hyper extremely bigger with C++ than with R. While I am actually more experienced with those languages than R, my ability to assist students that choose to do assignments in Python or C++ might be limited by time. Yet, I will absolutely always help find help material

Course (tentative) outline:

Date	Lecture Subject
01-11	Intro + Review
01-18	Review + Optimization
01-25	Intro to Comput. Game Theory -spatial models
02-01	Everyday data cleaning & handling
02-08	More advanced data handling
02-15	Web scraping - part 1
02-22	Web scraping - part 2
03-01	OLS regression: estimation, analysis and diagnostics
03-15	OLS regression: bootstrapping for hypothesis testing
03-22	Parallel computing: CPU multi-threading
03-29	Computational Game Theory - games*
04-05	Using the *ggplot* package for plotting*
04-12	Brief Intro to Text Analysis*
04-19	Brief Intro to Network Analysis*

* those are topics that are going to be covered only as time permit and are subject to change given demand.

Notice that the above outline is subject to changes. Three possible changes are particularly in the horizon:
(a) I may have to move around the two lectures on OLS, depending on how Polsci 699 goes. For obvious reasons I am interested in making it so that we see OLS regression topics just after or at the same time students are seeing in Polsci 699.
(b) I still have doubts whether we need two lectures for Web Scraping. I might make adjustments due to that.
(c) On 04-05 I may have to be in Chicago for the Midwest Political Science Association Annual Conference.

Grading:
The final grade will be composed of four parts:
- 10% will come from attendance and participation. That is, this is a **subjective** assessment.
- 40% will come from 4 homeworks.
- 15% will come from a student-made lecture note.
- 35% will come from a final project.

Date	Due assignments (always by 11:59pm)
01-26	homework 1
02-09	homework 2
02-23	homework 3
04-06	homework 4
04-13	student-made lecture notes
04-26	final project

Attendance and participation: if you are present every lecture, if you engage with the material, if you do not sleep in class and if you go talk to me when struggling with the material, then you are golden.

Homework: in the 4 homework you will be allowed to work either solo or with one colleague. However, you **cannot repeat the same colleague** in different homework. Groups greater than 2 will not be allowed under any circunstances. Talking to students other than your eventual pair is allowed and encouraged. However, if code is too similar to that of other pair
solo students, no grade will be given. The lowest homework grade will be dropped and there will be a late homework policy: if homework is submitted up to 5 hours late, no penalty will be given. From 5 hours up to 24 hours late, it will get a 10% discount (on top of raw grade). If homework is from 24 to 48 hours late, it will get a 20% discount. If homework is from 48 to 72 hours late, it will get a 30% discount. After 72 hours, no homework will be accepted.

Student-made lecture note: in consultation with the instructor, each student will choose an intermediate or advanced topic of interest, not covered in this class, and will prepare a lecture material (heavily commented code) as if they were going to teach a lecture on the topic. Grade will be assigned on the basis of challenge, clarity of comments and explanations, whether the code works without errors or bugs, as well as overall code quality. This is your opportunity to spend time learning how to learn by yourself and learning a topic (with guidance, since I will be always available to chat about it in office hours) that we do not have time to learn in class.

Final project: each student will submit a final project with at least 5 pages (single spaced), typeset in LaTeX, which should be a draft version of a real life analysis using programming knowledge learned in class or authorized by the instructor. Students can use whatever dataset they want and can address whatever substantive question they desire, but the final project absolutely must include at least two topics from the above Course Outline, and ***one of those two has to be*** either Computational Game Theory, Web Scraping, Paralel Computing or a special topic (more on this later). The choice of class topics has to be approved by the instructor. Now, the substantive question or theme to which the student will apply the techniques is, of course, totally free.

In all assignments, your grade will depend on (a) completeness (whether you did all the requested job); (b) functionality (whether your code has the functionality it should have); (c) accuracy (whether you get approximately the correct answeres from your implementations); (d) code quality (whether you wrote quality code instead of anything that simply gets the job done); (e) code clarity (how readable your code is). A really relevant part of your grade will depend on code quality and code clarity, so I strongly recommend taking a look at my lectures from Polsci 514 Fall on code quality. Also, check Google's R Style Guide.

Class Canvas:

The class entry in the university's Canvas system should be your main point of connection to this course. There you will find all the class material, there you will submit your homework, find your grades and other info. There you can also find a list of past announcements that I have sent to the group. Please do check the class' Canvas site fairly regularly.

Communication:

Strictly all regular communication will happen via Piazza (which means not via email), a very handy on-line questions-and-answers platform. In Piazza, students can ask questions anonymously, both publicly and privately (although the latter should be saved only for discussing personal matters or when the question being asked contains part of a solution for a homework). This way, we centralize our communication and it makes it so that all questions and doubts you might have end up reaching all colleagues. Another neat feature of Piazza is that students can also answer to students. If questions there become frequent (as I hope), I will include a grade bonus for those who also frequently answer colleagues' questions.

Disability Accommodations:

The Office of Services for Students with Disabilities (SSD) is located in G664 Haven Hall (763-3000, http://ssd.umich.edu/). SSD typically recommends accommodations through a Verified Individualized Services and Accommodations (VISA) form. Any information you provide is private and confidential and will be treated as such. If you need specific arrangements, please contact me directly via email rather soon.

Academic Integrity:

Here you find university's standards for academic and professional conduct:
http://www.rackham.umich.edu/current-students/policies/academic-policies/section11
Any breaches of academic integrity will be immediately reported both to the Political Science department and to university's relevant offices.