

POLSCI 514 Winter - Intermediate Programming Topics for Social Research

Instructor: Fabricio Vasselai (Winter 2021)

Friday 10am-11am at <https://umich.zoom.us/j/98678125624>

E-mail: vasselai@umich.edu

Office Hours: Wednesdays, 2pm-4pm, at <https://umich.zoom.us/j/98678125624>

Class at Canvas: <https://umich.instructure.com/courses/411475>

Class at Piazza: <http://piazza.com/umich/winter2021/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. The entirety of the class will be taught using R and its user-friendly IDE called RStudio. You should have both installed and available in all lectures. Students interested in learning additional programming languages (e.g. Python, C++) should come talk to me - I can absolutely help you with that.

If you took Polsci 514 Fall, note 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 around specific (often self-contained) topics. Accordingly, there will often be (pretty short) required readings, 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.

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.

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.

Note that the pace of this class will be significantly faster than the Fall version of the class, exactly because of the assumptions above and because we have half the time to meet.

Course (tentative) outline:

Lecture	Date	Lecture Subject
1	01/22	Review & Function Optimization
2	01/29	Basic Data Cleaning and Handling
3	02/05	More advanced data handling
4	02/12	Intro to Computational Game-Theory
5	02/19	Web-scraping - Part 1
6	02/26	Web-scraping - Part 2 *
7	03/05	How Computers Work & Code Performance
8	03/12	Parallel Computing
9	03/19	Special topic
10	03/26	Gaussian OLS regression
11	04/02	Gaussian OLS regression post-estimation and bootstrapping
12	04/09	Other regressions (Robust, Poisson, Negative Binomial, Logistic, Hierarchical)
13	04/16	Special topic

* I still have doubts whether we will use two lectures for Web Scraping. I might use that second day to teach another Special Topic.

Special topics

In those days, I will teach a 1-hour down-to-Earth overview of an applied topic of your choosing, naturally in R. Some options I am sure will be available for you to choose from: Intro to Computational Game Theory, Causal Inference, Intro to Text Analysis, Intro to Network Analysis, GIS data and spatial analysis. Some I am considering: Supervised Machine Learning, Unsupervised Machine Learning, Bayesian Models.

Grading:

The final grade will be composed of four parts:

- 10% will come from attendance and participation. That is, this is a **subjective** assessment.
- 50% will come from 4 homeworks.
- 20% will come from finishing weekly-ish (ungraded) DataCamp assignments in time.
- 20% will come from a final assignment (DataCamp + student-made lecture code)

Date	Due assignments (always by 11:59pm)
02-12	homework 1
03-05	homework 2
03-26	homework 3
04-16	homework 4
04-28	final assignment

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 must work alone. You are allowed to verbally discuss issues with your colleagues, but **you are not allowed to look at their code, neither entire code nor code parts**. After all, there is no such a thing as unseeing a code solution. If students' code is too similar to each other or to other sources, no grade will be given and **I will report to the department and to LSA**. 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 20% discount (on top of raw grade). If homework is from 24 to 48 hours late, it will get a 30% discount. If homework is from 48 to 72 hours late, it will get a 40% discount. After 72 hours, no homework will be accepted.

DataCamp weekly-ish assignments: most weeks, I will assign DataCamp course chapters on relevant topics for you to complete. While I will not grade them, they **must** be finished in time (no lateness allowed here) and with your best effort. DataCamp allows me to check on how you are progressing, for example if you

are just hitting the button that gives you free solutions nearly always. If I get students doing that, I will severely take marks from their grades.

Final assignment: each student will choose at least 4 DataCamp entire courses to complete, from a list I will offer. Those will be intermediate to somewhat advanced applied topics that we will not have the time to cover in this class. Then, each student also picks one of those topics to prepare their own lecture code on (as if they were teaching a lecture on the topic). I will grade such lecture codes and then **they will be shared with the whole class** (you will have the optional chance to review your code before sharing). 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.

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 answers 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.