Political Data Science (Beta)
L32 4625

Tuesdays and Thursdays
10:00-11:20AM
Seigle 104

Instructor Information

Jacob M. Montgomery, Ph.D.
Assistant Professor, Department of Political Science
Office: Seigle 285
E-mail: jacob.montgomery@wustl.edu
Telephone: None
Office Hours: Wed. 10-12 and by appointment

Course description

Rapid improvements in computing power, advances in powerful algorithms, and a massive increase in the availability of data has created a new world of possibilities for businesses, governments, policymakers, and decision-makers at every level. The goal of this class is to make you a participant rather than a spectator in this new data-driven world. In particular, the goal is to give you real-world skills and experience working with real data to answer real questions about our political world.

This class is designed to achieve four broad objectives. First, it aims to guide students as they learn the specifics of the R programming language, a powerful statistical computing environment widely used in the fields of political science, network analysis, machine learning, and statistics. Second, the class will also introduce students to some the basic approaches to data analysis, storage, and visualization in the machine learning/statistics literatures. So you are going to learn how to code. But more importantly, you will learn how to acquire, manipulate, store, visualize, and analyze data to answer real-world questions.

More broadly, this course aims to provide students some of the foundational concepts and skills needed to engage in modern data science. No course can teach you *everything* there is to know about R, machine learning, and statistical science even as it exists today. Even more certainly no class can teach you everything you will need to use in your future career. Some of the tools that will be in wide use in ten years do not even exist today. Thus, this course aims to give you the more foundational meta-skills you need to teach *yourself* the toolkit of the future. Learning at this level will also better equip you to understand results and analyses produced by others.

Finally, the end of the course will shift from the academic to the practica. Students will be tasked with tackling a real world political problem using the skills they have learned in the course. A major component of the course includes learning how to plan and execute a collaborative, complex data science project as well as how to effectively document and communicate the results.

Learning objectives

By the end of this course, you should be able to:

- Explain the basic components of the R working environment
- Understand object-oriented programming (or at least R's version of this)
- Understand the basic control functions, flow functions, and data structures of R
- Functionalize complex and/or repetitive code
- Clearly document code and develop a codebase on a collaborative platform
- Read-in and write-out data and text of any format, including information collected online
- Store complex data in a database
- Create custom data visualizations
- Fit basic machine learning models and be able to interpret the results
- Create a basic website and webapp
- Work collaboratively to plan and execute a question-driven data science project

Requirements and Evaluation

Grading in this class will be based on the components described below. **Late work will not be accepted without prior permission**. Makeup exams will not be given, and students who miss exams will receive a score of 0 absent extraordinary circumstances.

Grading scale

Score	Grade	Score	Grade	Score	Grade	Score	Grade
$\geq$94	A	$\geq$83	B	$\geq$ 73	C	$\geq$63	D
$\geq$90	A-	$\geq$80	B-	$\geq$ 70	C-	$\geq$60	D-
$\geq$87	B+	$\geq$77	C+	$\geq$ 67	D+	$<$60	Fail

Peer assessments - 10%

You will be assigned into a team of 3-5 individuals. You will work with this team throughout the semester on in-class assignments and your final research project. To help ensure that all members of the team are actively contributing, students will be asked to evaluate their teammates' contributions, effort, and performance. You will receive ungraded midterm evaluations from your team to help you know how well you are doing and identify areas in need of improvement.

Problem sets and in-class work - 30%

Problem sets, or homeworks, will be distributed throughout the course (20%). *Unless otherwise specified*, these are individual assignments that you should prepare yourself, though you may ask your colleagues for help. To be clear, *every single keystroke for these assignments should be your own*. DO NOT COPY AND PASTE CODE FROM YOUR PEERS, THE INTERNET OR ANYWHERE ELSE.

Please turn them in at the on the specified date **at the beginning of class**. If you have a printing problem, you are responsible for emailing it to me or the graduate TA before class starts. Each student's lowest homework grade will be dropped in the final grade calculations. This option should be reserved for illness, family emergencies, broken alarm clocks, or other unforeseen events. No additional waivers will be granted.

In-class assignments will be completed during class with your team (10%). All members will turn in a single assignment at the end of class and will share their grade. However, **all absent students will receive a zero**. Students missing more than five minutes of class time will be counted as absent. Each student's two lowest in-class assignment grades will be dropped in the final grade calculations. This option should be reserved for illness, family emergencies, broken alarm clocks, or other unforeseen events. No additional waivers will be granted.

Midterm exam - 25%

The midterm exam will be a take-home exam where you will be expected to independently create an organized codebase to accomplish a specific task. The exam will be due at the beginning of class on March 8. Specific rules will be explained at the time of the exam.

Project - 35%

- **Graduate Students**: After the midterm exam, teams will be assigned a specific programming task of interest either to myself or another faculty member in the department. Working with your assigned teams, and under the close supervision of the faculty member, students will be responsible for planning, creating, and documenting an analysis that meets the specified needs of the faculty member.
- **Undergraduates**: After the midterm exam, teams will consult with me to choose a project of their own for which they will generate a data science solution. Working with your assigned teams, students will be responsible for planning, creating, and documenting a codebase to achieve the goal as chosen by the team. This will include outlining the research goals, developing a plan, monitoring the progress of the team to specific benchmarks, and evaluating the final product.

The course will culminate with turning over the results in the form of a website or app to me for grading at the time of the regularly scheduled final for this course.

Grades are final

No adjustments will be made to final grades under any circumstances and no incompletes will be granted absent extraordinary circumstances.

Class policies

Teaching Assistant

There is one graduate teaching assistant. She will work closely in conjunction with Professor Montgomery on all issues of grading, but all grading decisions will be mine.

Dominique Lockett
DLOCKETT@WUSTL.EDU
Office Hours: Tuesdays, 1pm-3pm
Office location: Seigle 278

There is also one undergraduate teaching assistant. She will work on the online course book and hold office hours.

Mariah Yelenick
MYELENICK@WUSTL.EDU
Office Hours: Tuesdays, 4pm-6pm
Location: Whispers
(if you can't find her, message her on Slack)

Communications/Slack/Github

Email. is. just. the. worst. Using my data science skills I estimate that I get 1,385.3 emails/day (this is made up). For this course I have made a Slack workspace at `pds-class.slack.com`. Please install the App on your computer and/or phone and add this workspace. All class communications will occur here. We will also make individual channels for each problem set and groups are encouraged to make their own channels to facilitate communications.

You are, of course, allowed to communicate with me or the TA privately through Slack or email. However, unless it is confidential, I prefer that questions appear on the general channels so all students can see the question, the answer, and maybe offer solutions themselves.

Later in the semester we will also introduce you to Git and GitHub. GitHub is a powerful platform for collaborative and open data science and once you get the hang of it all problem sets, exams, and projects will occur here. It also offers nice solutions for hosting basic webpages. So go ahead and sign yourself up for a GitHub account now if you don't already have one.

Technology in the classroom

You will frequently make use of computers in this course. Plan to bring your laptop to class and have it ready to use each day. Please be respectful to your instructors and your peers by using your computers only for class-related purposes. Please put your phone away before class starts and don't bring it out. If I find you tweeting, playing games, watching sports, shopping, moderating Reddit, whatever TikTok is, or curating your Instagram feed, I will immediately stop class to have a discussion with you about what aspect of the current class session you find boring or uninteresting. You have been warned.

Academic Honesty

Cheating and plagiarism will not be tolerated. I strongly encourage you to review the University's policies regarding academic honesty, which you can read at: `http://www.wustl.edu/policies/undergraduate-academic-integrity.html`.

In general, if you have any question, please feel free to ask your TA or Professor Montgomery. Specific rules for this course:

- You may work together on homework in small groups, but you should each prepare your answers separately unless otherwise instructed.

- The homeworks and in-class work are "open book" and "open notes."
- You are to consult *only* with Professor Montgomery or a TA during exams.
- See the discussion of the rules for the problem sets above.

All cases of cheating or plagiarism will be referred to Washington University's Committee on Academic Integrity. If the Committee on Academic Integrity finds a student guilty of cheating, then the penalty will be (without exception) automatic failure of the course.

Students with disabilities

Students with disabilities enrolled in this course who may need disability-related classroom accommodations are encouraged to make an appointment to see me before the end of the second week of the semester. All conversations will remain confidential. Please also arrange to have the required documentation sent to me for any accommodations *at least two weeks prior to the first exam.*

Religious observances

Some students may wish to take part in religious observances that occur during this semester. If you have a religious observance that conflicts with your participation in the course, please meet with me *before the end of the second week of the semester* to discuss accommodations.

Reporting and accommodations for sexual assault

If a student discusses or discloses an instance of sexual assault, sex discrimination, sexual harassment, dating violence, domestic violence or stalking, or if a faculty member otherwise observes or becomes aware of such an allegation, they will keep the information as private as possible, but as a faculty member of Washington University, they are required to immediately report it to the Department Chair or Dean or directly to Ms. Jessica Kennedy, the University's Title IX Director, at (314) 935-3118, jwkennedy@wustl.edu. Additionally, you can report incidents or complaints to the Office of Student Conduct and Community Standards or by contacting WUPD at (314) 935-5555 or your local law enforcement agency.

The University is committed to offering reasonable academic accommodations (e.g., no contact order, course changes) to students who are victims of relationship or sexual violence, regardless of whether they seek criminal or disciplinary action. If you need to request such accommodations, please contact the Relationship and Sexual Violence Prevention Center (RSVP) at rsvpcenter@wustl.edu or 314-935-3445 to schedule an appointment with an RSVP confidential, licensed counselor. Information shared with counselors is confidential. However, requests for accommodations will be coordinated with the appropriate University administrators and faculty.

Course materials

Textbooks

In addition to assigned readings that will be posted on Blackboard, the following books are required. You may be able to find free versions online or through the library.

de Vries, Andrie and Joris Meys. 2015. *R for Dummies* (2nd Edition). Wiley.

Wickam, Hadley and Garrett Grolemund. *R for Data Science*. OReilly. `https://r4ds.had.co.nz/`

Wickham, Hadley. *Advanced R*. CRC Press. `https://adv-r.hadley.nz/`

Wickham, Hadley. *Mastering Shiny*. O'Reilly. `https://mastering-shiny.org/`

Baumer, Benjamin, Daniel K. Kaplan, and Nicholas J. Horton. *Modern Data Science with R*. CRC.

Nolan, Deborah and Duncan Temple Lang. *XML and Web Technologies for Data Sciences with R*. Springer.

Additional readings/links/resources provided at `politicaldatascience.com`

Software and hardware

You will be using the R statistical package (`http://www.r-project.org/`). While R is available for every computing platform, some of the more advanced tasks performed in this class will be taught based on the assumption that you are working on a Mac or Windows machine. Linux users will need to work with me to find solutions on your machine. Note: *Once you are set up with R, R-Studio, and Git, do NOT update your operating system (OS) for the remainder of the semester. If you are considering updating your OS, do it now*.

Plan of the course

The basic outline of the course is divided into four components. In Section 1, we will introduce the R computing environment and develop some basic and advanced skills and topics. In this portion of the class, we will begin with short lectures and discussions of the assigned readings. To encourage engagement with these materials, students will be asked to accomplish assigned programming tasks both inside and outside of the class period.

In Section 2, we will move onto making use of existing software packages that allow you to acquire, manipulate, and visualize data. Largely, this will come in the form of learning to use some of the functionality offered in the *tidyverse* of R packages.

In Section 3, we will introduce several basic approaches to analyzing data. The goal will be to provide an intuition about these models as well practical advice as to how to fit, interpret, and evaluate performance.

In Section 4, we will move from the abstract to the applied as the class takes on several real-world statistical programming challenges. Under the close supervision of me (, each team will work on a project requiring the development of a complex set of code to answer a real-world question about our political world. As part of this Section, several class periods will be dedicated to covering some advanced topics that will be needed to execute their projects. Likely topics include webscraping, API access, SQL databases, website creation/hosting, and Shiny. By the end of this Section, each team will be able to produce a complete data science project that will pull together skills in coding, data acquisition, data visualization, machine learning, and website/app development.

Very tentative Schedule

All told, this is an ambitious project that will require a substantial intellectual engagement from each student. It will also require *flexibility* since the course may evolve – perhaps substantially – during the semester in response to the needs of the project, our results, and issues raised by students. Y'all are coming from a lot of different backgrounds so I will need to adjust pacing and structure in response.

In particular, you will be expected to work with your teams throughout the semester both inside and outside the class. You will be involved in many collaborative projects in your career, so consider building a positive working dynamic within your team to be part of the assignment. I also expect that students be quick to inform me and/or the TA when assignments seem vague, overly difficult, confusing, or incomplete. **The schedule below should be viewed as no more than suggestive**.

Date	Topic	Reading	Assignments	Notes
Section 1				
1/14	‘‘Hello world’’	DVM Chpt 1-2, Appendix *Advanced R* Chapter 1		Pre-test
1/16	Calculation/Structures	DVM Chpt 4,5 *Advanced R* Chapter 2		PS 1 dist.
1/21	Structures 2	DVM Chpt 7 *Advanced R* Chpts. 3, 4		
1/23	Control/flow	DVM Chpt 9	PS 1 due	
1/28	Functions	DVM Chpt 8 *Advanced R* Chpt 6		Teams assigned
1/30	Version control Documentation	DVM Chpt 11 *Advanced R* Chapter 5, 27		
2/4	Debugging Testing/evaluating	*Advanced R* Chpt 22-24		
Section 2				
2/6	`ggplot`	*R4DS* Chpt 3	PS 2 due	
2/11	`dplyr`	*R4DS* Chpt 5		
2/13	`tidy`	*R4DS* Chpt 11-12		
2/18	Relational data	*R4DS* Chpt 13		
2/20	Text-as-data	*R4DS* Chpt 14 *MDSwR* Chpt 15	PS 3 due	Practice midterm dist.
2/25	pipes, `map`, `walk`	*R4DS* Chpt 18, 21		
2/27	Functional programming `purrr`	*Advanced R* Chpt 9-11		

Date	Topic	Reading	Assignments	Notes
Section 3				
3/3	Machine learning intro			Midterm distributed
3/5	Supervised learning 1	*MDSwR* Chpt 7	Midterms due	Midterm evals
3/17	Supervised learning 2	*MDSwR* Chpt 8		PS 4 dist.
3/19	Unsupervised learning 1	*MDSwR* Chpt 9.1		Also topic models
3/24	Unsupervised learning 2	*MDSwR* Chpt 9.2		
3/26	Causality 1: A/B testing	Online content	PS 4 Due	PS 5 Dist.
3/31	Causality 2 : Observational strategies	Online content		
Section 4				
4/2	Interactive visualizations	*Mastering Shiny* 1-5 *MDSwR* Chpt 11		
4/7	Webscraping 1	*XML* Chpt 1-3		
4/9	Webscraping 2	*XML* Chpt 4-5	PS 5 due	
4/14	SQL and friends	*MDSwR* Chpt 12, 13		
4/16	Github pages			Dom (MPSA)
4/21	Using APIs	*XML* 8,10		
4/23	TBD			Post-test
5/4	FINAL PROJECTS DUE			

Peer evaluation form (end of semester)

Name/team #:

Please assign scores that reflect how you really feel about the extent to which the other members of your team contributed to your learning and/or your teamÕs performance. This will be your only opportunity to reward the members of your team who worked hard on your behalf. (Note: If you give everyone pretty much the same score, you will be hurting those who did the most and helping those who did the least.)

Instructions: In the space below, please rate each of the other members of your team. Each member's peer evaluation score will be the average of the points they receive from the other members of the team. To complete the evaluation you should: 1) List the name of each member of your team in the alphabetical order of their last names and, 2) assign an average of ten points to the other members of your team and, 3) differentiate some in your ratings; for example, you must give at least one score of 11 or higher (maximum = 15) and one score of 9 or lower.

	Team member	**Score**
1.		
2.		
3.		
4.		

Additional feedback

Please briefly describe the reasons for your highest and lowest ratings in the space below. These comments will be shared anonymously. Note: Your comments should be descriptive, not evaluative; as clear and specific as possible; phrased in constructive terms; and focused on areas in which the student has made especially valuable contributions or could improve in the future.

Reason(s) for your highest rating(s):

Reason(s) for your lowest rating(s):