POLSCI 373 (INTSTD 401):
Data Science for International Studies

Christopher Fariss (cjfariss@umich.edu)

Office: Institute for Social Research, Atrium area

In-person (and Zoom) Office Hours: Tuesday 1:30am-5:30pm,
and by appointment.

Last Updated: January 9, 2025

Introduction

Course Content

Learn to program, program to learn. This class will provide undergraduate students with an introduction to the scientific method and programming tools for data science. Students will learn the fundamentals of the scientific method and, through programming and research design, how to improve both causal inference and the measurement of international political phenomena.

Course Objectives

Students will learn how to design studies to take advantage of the wealth of information contained in new online datasets such as data available from Social Media, newly digitized document corpuses now available online, and event-based datasets such as the ACLED (Armed Conflict Location and Event Data) dataset. The focus of the course is on building data science tools in such a way as to maximize the validity of inferences obtained from complex datasets like these.

- Learn to program, program to learn.

- Learn to program in R using the base package and a minimal number of other packages

- Collect and analyze international relations data using R

- Use the tools from research design to understand the what we can and cannot determine about the relationship between two or more variables

How will we go about learning these tools? In this class, we will learn to program and program to learn. What do I mean? First, we will use the R program environment to learn the building blocks of programming. As we develop programming skills in R, we will use them to help us understand how different types of data analysis tools work. For example, by the end of the course, students will be able to program and evaluate a model of data from several datasets of interest to students of world politics and international relations.

The primary learning tool in this course is a set of **Program Challenges**. I have designed a set of programming tools that students can then use to explore and understand the basic building blocks of any statistical model or data science question that students encounter. The program challenges are designed to be increasingly challenging, but also fun and collaborative. I push my students to think creatively, while also reminding them that getting stuck or not finishing all the steps in any program challenge is expected and encouraged, because it is a sign that they are learning. I have designed the program challenges and other programing and statistical material to be extensible so that students with limited skills all the way to advanced skills can learn from the material. I explain in each program challenge that students should always start with step 1 and then continue to each step as time permits. I also emphasize that the students should not worry about completing each step but that they may wish to come back to steps that they found difficult as they progress through the course. Each student will find different steps challenging and this is where the individualized learning takes place. These structural choices allow students from diverse educational backgrounds to learn and grow in my class.

Students learn in my course by finding the challenging steps in the Program Challenge that we use each week of the course. Students discover which steps in the Program Challenge are difficult for them and grow by tackling these steps. I encourage my students to show me where they are getting stuck so that I can help them make progress. Not finishing all the steps is expected and encouraged because the extensible structure of each Program Challenge means that any student can learn from making it through just a few steps. As students learn and gain confidence, they can return to the more difficult step in a Program Challenge from an earlier week in the course to learn and grow even more. This is an important requirement in my course, because at the midterm and final exam period, I ask my students to go back and complete additional difficult steps from their favorite program challenges.

Required Books

1. Davies, Tilman M. 2016. *The Book of R: A First Course in Programming and Statistics*. no starch press.

2. Matloff, Norman. 2011. *Art of R Programming: A Tour of Statistical Software Design*. no starch press.
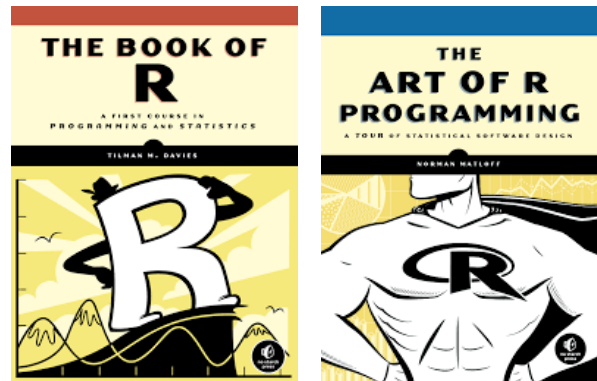
Required Articles

1. Cordell, Rebecca, K. Chad Clay, Christopher J. Fariss, Reed M. Wood, and Thorin M. Wright. 2022. "Recording Repression: Identifying Physical Integrity Rights Allegations in Annual Country Human Rights Reports" *International Studies Quarterly* 66(2):sqac016.

2. Geddes, Barbara. 1990. "How the Cases You Choose Affect the Answers You Get." *Political Analysis* 2:131-150. https://doi.org/10.1093/pan/2.1.131.

3. Hyde, Susan. 2015. "Experiments in International Relations: Lab, Survey, and Field." *Annual Review of Political Science* 18:403-424. https://doi.org/10.1146/annurev-polisci-020614-094854.

4. Raleigh, Clionadh, Andrew Linke, Håvard Hegre and Joakim Karlsen. 2010. "Introducing ACLED-Armed Conflict Location and Event Data" *Journal of Peace Research* 47(5):651-660. https://doi.org/10.1177%2F0022343310378914

5. Rubin, Donald B. 2008. "For Objective Causal Inference, Design Trumps Analysis." *Annals of Applied Statistics* 2(3):808-840. https://doi.org/10.1214/08-AOAS187

Web Documentaries and Lectures

We will also watch some short, 5-15 minutes web based documentaries and lectures about using data to understand trends around the globe.

1. Roser, Max. 2015. "Good Data will Make You an Economic Optimist"
   https://www.youtube.com/watch?v=519RSd65yFw

2. Rosling, Hans. 2006. "The best stats you've ever seen" (February 2006)
   https://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_
   ever_seen

3. Rosling, Hans. 2007. "New insights on poverty" (March 2007)
   https://www.ted.com/talks/hans_rosling_reveals_new_insights_on_poverty

4. TBD

5. TBD

6. TBD

Data Science Projects

1. Case selection essay

The Case selection essay is a 2-4 page essay (12-point font, 1-inch margins, double space). In consultation with the instructor, select a country in the world. The country needs to be included in the ACLED dataset: https://www.acleddata.com/wp-content/uploads/dlm_uploads/2018/02/Country-and-Time-Period-coverage_updatedJune2019.pdf.

In the essay, you should provide a brief summary that describes why you choose the specific country case. That is, explain why you are curious about the region. What about the country is interesting to you? What are the features about the country that led you to select it? You will select 2 variables from https://ourworldindata.org/. How does your country compare to others for the two variables you selected?

You will use the country you select when gathering data for the ACLED and Text data assignments, which are both described below.

- Seawright, Jason. 2016. "The Case for Selecting Cases That Are Deviant or Extreme on the Independent Variable" *Sociological Methods & Research* 45(3):493-525. https://doi.org/10.1177/0049124116643556.

2. ACLED data analysis project

In consultation with the instructor, develop a visualization and statistical analysis for the daily ACLED event data associated with your selected country for the last 2 years.

Additional readings suggested for this project include:

- Raleigh, Clionadh, Andrew Linke, Håvard Hegre and Joakim Karlsen. 2010. "Introducing ACLED-Armed Conflict Location and Event Data" *Journal of Peace Research* 47(5):651-660. https://doi.org/10.1177%2F0022343310378914

3. Human rights text data analysis project

In consultation with the instructor, select a text dataset associated with your country.

- Cordell, Rebecca, K. Chad Clay, Christopher J. Fariss, Reed M. Wood, and Thorin M. Wright. "Recording Repression: Identifying Physical Integrity Rights Allegations in Annual Country Human Rights Reports" *International Studies Quarterly* 66(2):sqac016.

Active Learning and Attendance

As students, we learn more when we actively engage with material instead of passively consuming it. This insight is supported by extensive research from college-level courses.[1] Because our course does not contain an in-class passive lecture (see Class Schedule Key below), participation and therefore attendance is critical for achieving our shared learning goals and is therefore mandatory. We will program in class together each week during our class sessions.

Monday lessons will be recorded via zoom and available online. If you cannot attend class in person, you are welcome to join class virtually via zoom or watch the recorded lesson at a later time. If you are sick, I want you to take the time to rest and recover fully. I will help you catch up with the course material. Wednesday lab sessions will focus on the program challenges. These are available to work on outside of class as well. Please feel free to use office hours to ask questions about these challenges as you work through them.

Chat GPT Policy

- My advise for this class: don't use Chat GPT or google search. But ...

- You may choose to use U-M GPT or another Chat GPT program in this course.

- The University of Michigan provides access to a U-M GPT tool, which you read more about here: https://its.umich.edu/computing/ai/gpt-in-depth

- **If you choose to use U-M GPT or another Chat GPT program, you must submit (1) the prompt used to generate the output, (2) the specific GPT program used, and (3) the date and time you entered the prompt.**

- You can read more about creating effective prompts here: https://teamdynamix.umich.edu/TDClient/30/Portal/KB/ArticleDet?ID=10680.

---

[1] See for example: Louis Deslauriers, Logan S. McCarty, Kelly Miller, Kristina Callaghan, Greg Kestin. 2019. "Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom" *Proceedings of the National Academy of Sciences* 116 (39) 19251-19257; https://doi.org/10.1073/pnas.1821936116

Grade Values

Please note that it is possible to earn more than 100 points in this course.

Participation/Attendance: 40 points (20 points for the first half the semester and 20 points for the second half of the semester)

Case selection essay: 5 points

Student selected program challenges: 5 points each (max to 25 points)

ACLED data analysis challenge: 5 points

Text data analysis challenge: 5 points

Midterm Exam: 15 points

Final Exam: 15 points

Out of class video surveys: 1 point each (max 5 points)

Grade Ranges

A+ [97+] Please note that this option is only possible with additional projects and in consultation with the instructor.

A [93–96]

A- [90–93)

B+ [87–90)

B [83–87)

B- [80–83)

C+ [77–80)

C [73–77)

C- [70–73)

D [60–70)

F [0–60)

Additional Information about Office Hours

The best way to contact me in order to ask additional questions about course material is to attend office hours. I hold extensive office hours each week online on zoom and in-person in the atrium area on the first floor of the Institute for Social Research (ISR) building, which is located at 426 Thompson St., Ann Arbor, MI 48104. Just like in the class room, office hours sessions are designed to be an open and inclusive learning environment for my students. The ISR atrium offers an open space that accommodates small to medium size groups. There will often be more than one student seeking feedback during office hours. Each of you are welcome to join the group office hours session and ask questions or participate in discussion. I am also available outside of office hours via a message on the course management software (Canvas) or ay my umich email address: `cjfariss@umich.edu`.

Class Schedule

Class Schedule Key

Programming Lessons: An hour of class time will be devoted to programming in R together. Please make sure to read the assigned chapters and view the video lectures before each weeks classes so that you are ready to program in class. These lessons will occur on Mondays throughout the semester. These weekly lessons will review key information from the weekly reading assignments. Lessons will also be recorded on zoom and will be available online.

Program challenges: These challenges are meant to be just that, challenging. They should also be fun. I encourage you to think creatively and collaboratively. Getting stuck or not finishing all the steps is expected and encouraged. This is how learning works. Always start with step (1) and then continue to each step as time permits. Don't worry about completing each step. Document your code for each step. You may wish to come back to some of the harder steps as you progress through the course. Note that some of the steps may ask you to use skills we have not yet covered in the course. Don't worry about these steps now but definitely think through the programming logic if you are stuck and make plans to come back to try them once you feel ready. The program challenges will occur on Wednesdays throughout the semester.

In class reading discussion: Occasionally over the course of the term, we will read and discuss an academic article. Please read these articles prior to the class date.

Offline video: We will watch a few short videos about data science in some classes.

Readings for the week: Assigned readings (book chapters or articles). Please make sure to read these articles prior to the Wednesday class each week.

Week 1: Introduction

Readings for the Week:

1. Davies Ch.1 "Getting Started"

Monday — 01/06/2025

No class

Wednesday — 01/08/2025

- **Programming lesson:** 3 sets of 3

- **Programming lesson:** Download and install R and Rstudio

Week 2: Numbers in R and other stuff in R too

Readings for the Week:

1. Davies Ch.2 "Numerics, Arithmetic, Assignment, and Vectors

2. Matloff Ch.2 "Vectors" (optional)

3. Davies Ch.3 "Matrices and Arrays"

4. Matloff Ch.4 "Matrices and Arrays" (optional)

Monday — 01/13/2025

- **Programming lesson:** Introduction to programming in R

- **Programming lesson:** Making and manipulating matrices and arrays in R

Wednesday — 01/15/2025

- **Program challenge**

- **Out of class video:** Rosling, Hans. 2006. "The best stats you've ever seen" (February 2006) https://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen

Week 3: Case Selection

Readings for the Week:

1. Geddes (1990)

Monday — 01/20/2025

- No class in observance of Martin Luther King day.

Wednesday — 01/22/2025

- **Program challenge**

- **Out of class video:** Roser, Max. 2015. "Good Data will Make You an Economic Optimist"
  https://www.youtube.com/watch?v=519RSd65yFw

- **Programming lesson:** Making and manipulating matrices and arrays in R (continued)

- **Reading discussion:** Geddes, Barbara. 1990. "How the Cases You Choose Affect the Answers You Get." *Political Analysis* 2:131-150.

Week 4: Data inside and outside of R

Reminder: Case selection choice is due on the Canvas website.

Readings for the Week:

1. Davies Ch.8 "Lists and Data Frames"

2. Matloff Ch.4 "Lists" (optional)

3. Matloff Ch.5 "Dataframes" (optional)

4. Matloff Ch.6 "Factors and Tables" (optional)

Monday — 01/27/2025

- **Programming lesson:** working with lists, dataframes, and tables in R

Wednesday — 01/29/2025

- **Program challenge**

Week 5: Programming in R

Readings for the Week:

1. Davies Ch.10 "Conditions and Loops"

2. Matloff Ch.7 "R Programming Structures" (optional)

3. Matloff Ch.4 "Lists"

4. Matloff Ch.5 "Dataframes"

5. Matloff Ch.6 "Factors and Tables"

Monday — 02/03/2025

- **Programming lesson:** Conditions and Loops in R

- **Programming lesson:** working with lists, dataframes, and tables in R (continued)

Wednesday — 02/05/2025

- **Program challenge**

Week 6: Programming in R

Readings for the Week:

1. Davies Ch.9 "Calling Functions"

2. Davies Ch.11 "Writing Functions"

3. Davies Ch.8 "Reading and Writing Files"

4. Matloff Ch.10 "Input/Output" (optional)

Monday — 02/10/2025

- **Programming lesson:** Using and Writing functions in R

Wednesday — 02/12/2025

- **Program challenge**

Week 7: Programming in R

Readings for the Week:

1. Davies Ch.12 "Exceptions, Timing, and Visibility"

2. Matloff Ch.13 "Debugging"

Monday — 02/17/2025

- **Programming lesson:** Tips and tricks for dealing with errors in R

- **Programming lesson:** R programming review

Wednesday — 02/19/2025

- **Program challenge**

Week 8: Review

no readings this week

Monday — 02/24/2025

- **Midterm exam:** The R programming exam consists of two parts (1) an open book take home exam, and (2) and in-class program challenge. Each student should select one of the program challenges from the course so far and work to finish as many of the steps for the program challenge as possible. During the in-class portion, students may work collaboratively.

Wednesday — 02/26/2025

- **Programming lesson:** R programming challenge (completed 2 challenges to turn in as part of the midterm)

Spring Break!

Week 9: Statistical and Visual Summaries of Data

Readings for the Week:

1. Davies Ch.15 "Probability"

2. Davies Ch.16 "Common Probability Distributions"

3. Davies Ch.14 "Basic Data Visualization"

4. Matloff Ch.12 "Graphics" (optional)

Monday — 03/10/2025

- **Programming lesson:** Visualizing probability distributions in R

Wednesday — 03/12/2025

- **Program challenge**

Week 10: Statistical and Visual Summaries of Data

Readings for the Week:

1. Davies Ch.13 "Elementary Statistics"

2. Davies Ch.17 "Sampling Distributions and Confidence"

3. Raleigh et al. (2010)

Monday — 03/17/2025

- **Reading discussion:** Raleigh, Clionadh, Andrew Linke, Håvard Hegre and Joakim Karlsen. 2010. "Introducing ACLED-Armed Conflict Location and Event Data" *Journal of Peace Research* 47(5):651-660.

- **Programming lesson:** Visualizing and summarizing event count data from ACLED

Wednesday — 03/19/2025

- **Program challenge**

Week 11: Using Strings of Text as Data

Readings for the Week:

1. Cordell et al (2022).

2. Matloff Ch.11 "String Manipulation"

Monday — 03/24/2025

- Cordell, Rebecca, K. Chad Clay, Christopher J. Fariss, Reed M. Wood, and Thorin M. Wright. 2022. "Recording Repression: Identifying Physical Integrity Rights Allegations in Annual Country Human Rights Reports" *International Studies Quarterly* 66(2):sqac016.

- **Programming lesson:** Text data in R

Wednesday — 03/26/2025

- **Program challenge** using ACLED data and Text data

Week 12: Research Design and Data Analysis

Readings for the Week:

1. Davies Ch. 18 "Hypothesis Testing"

2. Davies Ch. 20 "Simple Linear Regression"

Monday — 03/31/2025

- **Programming lesson:** Programming statistical comparisons in R

- **Programming lesson:** Simple linear model in R

Wednesday — 04/02/2025

- **Program challenge**

Week 13: Research Design and Data Analysis

Readings for the Week:

1. Hyde (2015)

2. Rubin (2008)

3. Matloff Ch.8 "Doing Math and Simulations in R" (optional)

Monday — 04/07/2025

- **Reading discussion:** Hyde, Susan. 2015. "Experiments in International Relations: Lab, Survey, and Field." *Annual Review of Political Science* 18:403-424.

- **Reading discussion:** Rubin, Donald B. 2008. "For Objective Causal Inference, Design Trumps Analysis." *Annals of Applied Statistics* 2(3):808-840.

- **Programming lesson:** Simulating potential outcomes in R

Wednesday — 04/09/2025

- **Program challenge**

Week 14: Graphics

Due Date: Selected data visualization (ACLED or Text data) is due in class this week.

Readings for the Week:

1. Davies Ch.23 "Advanced Plot Customization"

2. Davies Ch.24 "Going Further with the Grammar of Graphics" (optional)

3. Davies Ch.25 "Defining Colors and Plotting in Higher Dimension" (optional)

Monday — 04/14/2025

- **Programming lesson:** Advanced graphics in R

- Data Visualization Critique

Wednesday — 04/16/2025

- **Program challenge**

Week 15: Next Steps

*no readings this week*

Monday — 04/21/2025

- Review session for final exam.

- **Out of class video:** Rosling, Hans. 2007. "New insights on poverty" (March 2007)
  https://www.ted.com/talks/hans_rosling_reveals_new_insights_on_poverty

Wednesday — 04/23/2025

- Finals study day (not an official class day)

- I will hold extra office hours today to answer any questions before the final exam

Finals Week

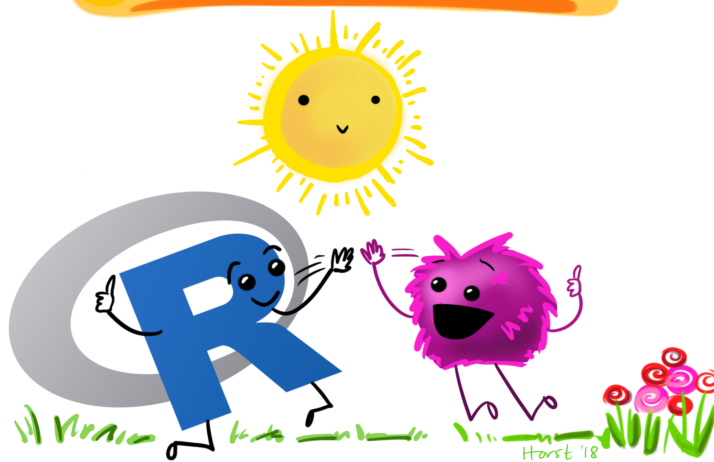- In-class portion of the final exam is scheduled for 04/30/2025: 4:00 pm - 6:00 pm
  https://ro.umich.edu/calendars/final-exams/winter-2025/
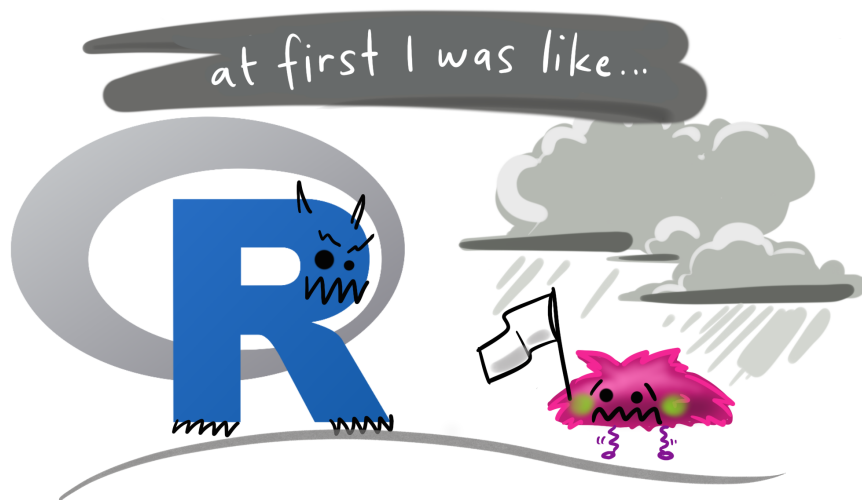
Figure 1: Artwork by @allison_horst

https://allisonhorst.com/

Additional Course Information

Student Mental Health and Wellbeing

University of Michigan is committed to advancing the mental health and wellbeing of its students. If you or someone you know is feeling overwhelmed, depressed, and/or in need of support, services are available.

For help, contact Counseling and Psychological Services (CAPS) at (734) 764-8312 and `https://caps.umich.edu/` during and after hours, on weekends and holidays, or through its counselors physically located in schools on both North and Central Campus.

You may also consult University Health Service (UHS) at (734) 764-8320 and `https://www.uhs.umich.edu/mentalhealthsvcs`, or for alcohol or drug concerns, see `www.uhs.umich.edu/aodresources`.

For a listing of other mental health resources available on and off campus, visit: `http://umich.edu/ mhealth/`.

Accommodations for Students with Disabilities

If you think you need an accommodation for a disability, please let me know at your earliest convenience. Some aspects of this course, the assignments, the in-class activities, and the way the course is usually taught may be modified to facilitate your participation and progress. As soon as you make me aware of your needs, we can work with the Services for Students with Disabilities (SSD) office to help us determine appropriate academic accommodations. SSD (734-763-3000; `http://ssd.umich.edu`) 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.

Religious and Academic Conflicts

Although the University of Michigan, as an institution, does not observe religious holidays, it has long been the University's policy that every reasonable effort should be made to help students avoid negative academic consequences when their religious obligations conflict with academic requirements. Absence from classes or examinations for religious reasons does not relieve students from responsibility for any part of the course work required during the period of absence. Students who expect to miss classes, examinations, or other assignments as a consequence of their religious observance shall be provided with a reasonable alternative opportunity to complete such academic responsibilities.

It is the obligation of students to provide faculty with reasonable notice of the dates of religious holidays on which they will be absent. Such notice must be given by the drop/add deadline of the given term. Students who are absent on days of examinations or class assignments shall be offered an opportunity to make up the work, without penalty, unless it can be demonstrated that a make-up opportunity would interfere unreasonably with the delivery of the course. Should disagreement arise over any aspect of this policy, the parties involved should contact the Director of Undergraduate Studies/Director of Graduate Studies. Final appeals will be resolved by the Provost.

Students Representing the University of Michigan

There may be instances when students must miss class due to their commitment to officially represent the University. These students may be involved in the performing arts, scientific or artistic endeavors, or intercollegiate athletics. Absence from classes while representing the University does not relieve students from responsibility for any part of the course missed during the period of absence. Students should provide reasonable notice for dates of anticipated absences and submit an individualized class excuse form.

Academic Integrity

The LSA undergraduate academic community, like all communities, functions best when its members treat one another with honesty, fairness, respect, and trust. The College holds all members of its community to high standards of scholarship and integrity. To accomplish its mission of providing an optimal educational environment and developing leaders of society, the College promotes the assumption of personal responsibility and integrity and prohibits all forms of academic dishonesty and misconduct. Academic dishonesty may be understood as any action or attempted action that may result in creating an unfair academic advantage for oneself or an unfair academic advantage or disadvantage for any other member or members of the academic community. Conduct, without regard to motive, that violates the academic integrity and ethical standards of the College community cannot be tolerated. The College seeks vigorously to achieve compliance with its community standards of academic integrity. Violations of the standards will not be tolerated and will result in serious consequences and disciplinary action.

Grade Grievances

If you believe a grade you have received is unfair or in error, you will need to do the following:
Wait 24 hours after receiving the grade before approaching your instructor. Provide an explanation in writing for why the grade you received was unfair or in error. If you believe the instructor's response fails to address your claim of unfairness or error, you may petition the department's Director of Undergraduate Studies at the latest within the first five weeks of classes following the completion of the course. You must convey in writing the basis for the complaint, with specific evidence in support of the argument that the grade either was given in error or was unfairly determined. This formal complaint also should summarize the outcome of the initial inquiry to the course instructor, indicating which aspects are in dispute. Within three weeks of the receipt of the petition, the DUS will determine whether to convene the Undergraduate Affairs Committee, the student, and the instructor(s) for a formal hearing. Further details on this process are included on the department website under Advising > Contesting a Grade.

Resources for Harassment

Title IX makes it clear that violence and harassment based on sex and gender, including violence and harassment based on sexual orientation, are a Civil Rights offense subject to the same kinds of accountability and the same kinds of support applied to offenses against other protected categories such as race, national origin, etc. If you or someone you know has been harassed or assaulted, you can find the appropriate resources here: www.bw.edu/resources/hr/harass/policy.pdf. For information about help and resources at University of Michigan please contact the Sexual Assault Prevention and Awareness Center (SAPAC) https://sapac.umich.edu/SupportServices; 734-764-7771; or sapac@umich.edu. For information about filing a report or complaint with the Title IX office at the University of Michigan see

Please note that Title IX offices often distinguish between making a "report," which does not launch an investigation, and filing a "complaint," which does.

Language and Gender

"Language is gender-inclusive and non-sexist when we use words that affirm and respect how people describe, express, and experience their gender. Just as sexist language excludes women's experiences, non-gender-inclusive language excludes the experiences of individuals whose identities may not fit the gender binary, and/or who may not identify with the sex they were assigned at birth. Identities including trans, intersex, and genderqueer reflect personal descriptions, expressions, and experiences. Gender-inclusive/non-sexist language acknowledges people of any gender (for example, first year student versus freshman, chair versus chairman, humankind versus mankind, etc.). It also affirms non-binary gender identifications, and recognizes the difference between biological sex and gender expression. Teachers and students should use gender-inclusive words and language whenever possible in the classroom and in writing. *Students, faculty, and staff may share their preferred pronouns and names, either to the class or privately to the professor, and these gender identities and gender expressions should be honored.*" For more information:
www.wstudies.pitt.edu/faculty/gender-inclusivenon-sexist-language-syllabi-statement.