Applied Data Science for Civil and Environmental Engineers

CGN 4905/6905

Class Periods: Tuesdays 10:40-11:30 am, Thursdays 10:40 – 12:30 pm

Location: Turlington Hall B310

Academic Term: Fall 2021

Instructor:
Dr. Xiang 'Jacob' Yan
xiangyan@ufl.edu
Office: Weil Hall 508

Office Hours: **Thursdays 1:30-3:00 pm**. You can either meet me in my office or on Zoom:
https://ufl.zoom.us/j/98439390395?pwd=YTN5T0k1V3Fadzd3Rm5UV2RFbW44Zz09
(Zoom Meeting ID: 984 3939 0395; Passcode: 718247)

If you cannot make it to the regular office hours, please email me for an appointment.

Course Description

Organizations in various industries and sectors are sparing no effort to extract actionable insights from big data. To help students learn how to use data to solve real-world problems, this course introduces them to the workflows of data science applications, covering the state-of-art techniques in data acquisition, data processing and management, analytics and modeling, and visualization. In addition to train students with fundamental knowledge of data science concepts and techniques, the course emphasizes the skills of critically applying such knowledge to complex socioeconomic and environmental contexts. Students will learn the basics of problem formulation and some major ethical considerations of applying data science in practice.

R will be the primary programming language taught in this class. However, students are allowed to finish the assignments using a different language they prefer, such as Python and Matlab. Students who choose to do so will have to work out the scripts on their own because sample codes and homework solutions will only be provided in R.

Course Pre-Requisites / Co-Requisites

Intro-level knowledge of programming languages such R, Python, and Matlab is highly recommended but not required. Students who have a strong interest but have no exposure to programming languages should consult the instructor. Strong math background is NOT required.

Course Objectives

Students are expected to learn the following from the class:

- Develop a basic understanding of various data-science techniques and skills
- Demonstrate an ability to define and formulate data-science problems
- Gain practical R programming skills and apply them to problem solving
- Communicate data insights effective through visualization tools
- Demonstrate knowledge of the main ethical considerations of data science practice

Required Textbooks and Software

None. All course materials will be uploaded on Canvas. Some class sessions have **required readings**, which I have uploaded onto Canvas, organized in folders by week/date. I have created a folder that contains some helpful optional readings that you may want to explore.

Course Schedule

Week	Lectures	Labs
1-Aug 24, Aug 26	Introduction to applied data science	Introduction to R/R Studio; Data import, export; data types
2-Aug 31, Sep 2	The potential and pitfalls of big data *(Short reading response 1 due before Aug 31 class)*	Basic operations in R and getting to know *Tidyverse*
3-Sep 7, Sep 9	Dates and times in R, the *lubridate* package *(Assign #1 due on Sep 7: R basics)*	Data wrangling with *dplyr;* descriptive analysis
4-Sep 14, Sep 16	Visualization *(Short reading response 2 due before Sep 14 class)*	R Markdown/Notebook, plotting with *ggplot2*, making interactive plots with *plotly*
5-Sep 21, Sep 23	Working with census data	Data transformation with *tidy* *(Assign #2 due on Sep 23: Descriptive analysis and plotting)*
6-Sep 28, Sep 30	Web-scrapping	Looping and APIs
7-Oct 5, Oct 7	Data joins	*No class (watch recorded video on data visualization)* *(Assignment #3 due on Oct 7: APIs, data transformation and data visualization)*
8-Oct 12, Oct 14	Introduction to spatial data and analytics	Making maps with *ggplot2*, *tmap*, *leaflet*
9-Oct 19, Oct 21	Spatial relationships and operations *(Assignment #4 due on Oct 19: Joining multiple datasets and making maps)*	Geospatial analysis with *sp* and *sf*; geometric operations
10-Oct 26, Oct 28	Statistical modeling *(Mini project proposal due)*	Interpreting model outputs (e.g., the *ilm* package)
11-Nov 2, Nov 4	Application #1: Infrastructure planning	Going through R codes for Application #1
12-Nov 9, Nov 11	Application #2: Flooding mapping *(Assignment #5 due on Nov 8: Spatial analysis and visualization; linear regression)*	*No Class (Holiday)*
13-Nov 16, Nov 18	Going through R codes for Application #2 (*Mini project report due on Nov 16*)	Guest lecture from a practitioner Application #3: Environmental justice
14-Nov 23, Nov 25	Going through R codes for Application #3 *(Final project proposal due on Nov 23)*	Help students with final project
15-Nov 30, Dec 2	Data ethics, privacy, and civic engagement *(Short reading response 3 due before Nov 30 class)*	Help students with final project
16-Dec 7	Final project presentations (or a poster fair of final project demonstrations)	(*Final project report due on December 14*)

Useful FREE E-books:

1. Peng, R. (2012). R Programming for Data Science [very introductory book]. Available at =
2. Wickham, H., Grolemund, G. (2021). R for Data Science. Available at https://r4ds.had.co.nz/data-visualisation.html
3. Wilke, C. Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures. Available at https://clauswilke.com/dataviz/index.html
3. Lovelace, R., Nowosad, J., Muenchow, J. (2021). Geocomputation with R. Available at https://geocompr.robinlovelace.net/index.html
4.Pebesma, E., Bivand, R. (2021). Spatial Data Science with Applications in R. Available at https://keen-swartz-3146c4.netlify.app/

Attendance Policy

Due to COVID 19, students can either attend either the face-to-face sessions or watch the Zoom recording. See UF policies on attendance at: https://catalog.ufl.edu/UGRD/academic-regulations/attendance-policies/

Evaluation of Grades

Assignment	Percentage of final grade
Assignment 1	10
Assignment 2	10
Assignment 3	10
Assignment 4	10
Assignment 5	10
Mini project proposal	5
Mini project report	10
Final project proposal	7
Final project poster	15
Final project report	8
Class attendance* (Reading responses)	5

Note:* Due to COVID-19, you can either attend classes in person or by watching the Zoom recordings. Performance of class attendance will mainly be evaluated based on your responses to the assigned readings. These responses can be very short (several sentences are good enough). You may write down a few key takeaways, raise some questions, or provide a short critique.

All assignments will have a strict deadline that you are expected to follow. I will only grant exceptions under extraordinary circumstances (please let me be aware of it before the assignment deadline). If you have missed the deadline, 20% of the full grade will be deducted from your final grade for each day that passes the deadline. For instance, if the deadline is October 1st and the full grade is 20, 4 points will be taken off if you submit on October 2nd, 8 points on October 3rd,… and you will get 0 point if you submit on October 6th or later.

Mini Project

The mini project is designed to help you be better prepared for the final project. In any data science project, the workflow usually goes as follows:

- Step 1: Define the problem that you need to solve with data.
- Step 2: Collect the data needed to solve the problem.
- Step 3: Clean, process, and analyze the data
- Step 4: Present the results with visualization (tables, graphs, and maps) and summarize main findings.

In this mini project, you are asked to do a small-scale project by completing all four steps.

Mini project proposal (11 fond size, 1.5 space, max 1 pages): In this proposal, state the problem you intend to address, list the specific questions that you will answer, describe the dataset(s) you intend to use, and discuss the analysis that you will conduct and the techniques (R packages) you envision to use. The key to a successful mini project is to only ask questions that leads to a reasonable project scope.

Problem: Is there an urban/rural disparity in the impacts of COVID-19 on US communities

Manageable research questions:

- Are there more COVID cases/deaths per capita in rural counties than in urban counties?

Unmanageable or unachievable research questions:

- Are there more COVID cases/deaths per capita in rural communities than in urban communities?
  (Reason: the data available to us is only available at the county level)
- If there are disparities, what factors cause them?
  (Reason: addressing this question requires collecting data on a lot of potential contributing factors and to do sophisticated statistical modeling, which is far beyond the scope of a course project)

Mini project report: The final report should be a HTML or PDF file generated by R Markdown detailing the problem statement, data collection, analysis, and results

Final Project
The final project will be a mini data science project done in teams of 2-3 people (single person or four person teams will be considered on an individual basis). It will have three deliverables: a proposal, a poster, and a final report.

Project proposal (11 fond size, 1.5 space, max 2 pages): I have assembled a set of datasets provided by public agencies. These datasets cover topics in public transportation, traffic volumes, crash incidents, food systems, soil quality, and hurricane/flooding. Feel free to choose a dataset that you have access to on your own. If you choose to work on the datasets that I provide, you may have a chance to work with the data provider (ask me to facilitate if you want to do this) and propose a project of interest to their agency/organization. The proposal should include:

1. A statement of the problem you intend to address. The problem should lead to a reasonable project scope, and the problem should be solvable with available data. In the problem statement, list the specific questions that you intend to answer in this project.
2. The dataset(s) you intend to use
3. A description of the analysis that you will conduct and the techniques you envision to use.
4. The visualization you plan to deliver in the final product. What types of visualization (charts, graphs, maps, tables)? Will you develop interactive features and/or animations? What are the main messages do you plan to convey through these visualizations?
5. A work plan that shows a breakdown of tasks, key dates, and key milestone.

Final project poster: You will prepare a poster to present your project to the public audience. The size should be size should be 24 $\times$ 36 inches.

Final project report: The final report should be a HTML or PDF file generated by R Markdown detailing the problem statement, data collection, analysis, and results. Include the R scripts in the report and try your best to make it visually appearing. See some good examples of final report written by R Markdown here: https://pennmusa.github.io/MUSA_801.io/

Grading Policy
The following is given as an example only.

Percent	**Grade**	**Grade Points**
93.4 - 100	A	4.00
90.0 - 93.3	A-	3.67
86.7 - 89.9	B+	3.33
83.4 - 86.6	B	3.00
80.0 - 83.3	B-	2.67
76.7 - 79.9	C+	2.33
73.4 - 76.6	C	2.00
70.0 - 73.3	C-	1.67
66.7 - 69.9	D+	1.33
63.4 - 66.6	D	1.00
60.0 - 63.3	D-	0.67
0 - 59.9	E	0.00

More information on UF grading policy may be found at:
http://gradcatalog.ufl.edu/content.php?catoid=10&navoid=2020#grades

Students Requiring Accommodations
Students with disabilities who experience learning barriers and would like to request academic accommodations should connect with the disability Resource Center by visiting https://disability.ufl.edu/students/get-started/. It is

important for students to share their accommodation letter with their instructor and discuss their access needs, as early as possible in the semester.

Course Evaluation
Students are expected to provide professional and respectful feedback on the quality of instruction in this course by completing course evaluations online via GatorEvals. Guidance on how to give feedback in a professional and respectful manner is available at https://gatorevals.aa.ufl.edu/students/. Students will be notified when the evaluation period opens, and can complete evaluations through the email they receive from GatorEvals, in their Canvas course menu under GatorEvals, or via https://ufl.bluera.com/ufl/. Summaries of course evaluation results are available to students at https://gatorevals.aa.ufl.edu/public-results/.

In-Class Recording
Students are allowed to record video or audio of class lectures. However, the purposes for which these recordings may be used are strictly controlled. The only allowable purposes are (1) for personal educational use, (2) in connection with a complaint to the university, or (3) as evidence in, or in preparation for, a criminal or civil proceeding. All other purposes are prohibited. Specifically, students may not publish recorded lectures without the written consent of the instructor.

A "class lecture" is an educational presentation intended to inform or teach enrolled students about a particular subject, including any instructor-led discussions that form part of the presentation, and delivered by any instructor hired or appointed by the University, or by a guest instructor, as part of a University of Florida course. A class lecture does not include lab sessions, student presentations, clinical presentations such as patient history, academic exercises involving solely student participation, assessments (quizzes, tests, exams), field trips, private conversations between students in the class or between a student and the faculty or lecturer during a class session.

Publication without permission of the instructor is prohibited. To "publish" means to share, transmit, circulate, distribute, or provide access to a recording, regardless of format or medium, to another person (or persons), including but not limited to another student within the same class section. Additionally, a recording, or transcript of a recording, is considered published if it is posted on or uploaded to, in whole or in part, any media platform, including but not limited to social media, book, magazine, newspaper, leaflet, or third party note/tutoring services. A student who publishes a recording without written consent may be subject to a civil cause of action instituted by a person injured by the publication and/or discipline under UF Regulation 4.040 Student Honor Code and Student Conduct Code.

University Honesty Policy
UF students are bound by The Honor Pledge which states, "We, the members of the University of Florida community, pledge to hold ourselves and our peers to the highest standards of honor and integrity by abiding by the Honor Code. On all work submitted for credit by students at the University of Florida, the following pledge is either required or implied: "On my honor, I have neither given nor received unauthorized aid in doing this assignment." The Conduct Code (https://sccr.dso.ufl.edu/process/student-conduct-code/) specifies a number of behaviors that are in violation of this code and the possible sanctions. If you have any questions or concerns, please consult with the instructor or TAs in this class.

Commitment to a Safe and Inclusive Learning Environment
The Herbert Wertheim College of Engineering values broad diversity within our community and is committed to individual and group empowerment, inclusion, and the elimination of discrimination. It is expected that every person in this class will treat one another with dignity and respect regardless of gender, sexuality, disability, age, socioeconomic status, ethnicity, race, and culture.

If you feel like your performance in class is being impacted by discrimination or harassment of any kind, please contact your instructor or any of the following:
• Your academic advisor or Graduate Program Coordinator
• Jennifer Nappo, Director of Human Resources, 352-392-0904, jpennacc@ufl.edu

• Curtis Taylor, Associate Dean of Student Affairs, 352-392-2177, taylor@eng.ufl.edu
• Toshikazu Nishida, Associate Dean of Academic Affairs, 352-392-0943, nishida@eng.ufl.edu

Software Use

All faculty, staff, and students of the University are required and expected to obey the laws and legal agreements governing software use. Failure to do so can lead to monetary damages and/or criminal penalties for the individual violator. Because such violations are also against University policies and rules, disciplinary action will be taken as appropriate. We, the members of the University of Florida community, pledge to uphold ourselves and our peers to the highest standards of honesty and integrity.

Student Privacy

There are federal laws protecting your privacy with regards to grades earned in courses and on individual assignments. For more information, please see: https://registrar.ufl.edu/ferpa.html

Campus Resources:

Health and Wellness

U Matter, We Care:

Your well-being is important to the University of Florida. The U Matter, We Care initiative is committed to creating a culture of care on our campus by encouraging members of our community to look out for one another and to reach out for help if a member of our community is in need. If you or a friend is in distress, please contact umatter@ufl.edu so that the U Matter, We Care Team can reach out to the student in distress. A nighttime and weekend crisis counselor is available by phone at 352-392-1575. The U Matter, We Care Team can help connect students to the many other helping resources available including, but not limited to, Victim Advocates, Housing staff, and the Counseling and Wellness Center. Please remember that asking for help is a sign of strength. In case of emergency, call 9-1-1.

Counseling and Wellness Center: https://counseling.ufl.edu, and 392-1575; and the University Police Department: 392-1111 or 9-1-1 for emergencies.

Sexual Discrimination, Harassment, Assault, or Violence

If you or a friend has been subjected to sexual discrimination, sexual harassment, sexual assault, or violence contact the **Office of Title IX Compliance**, located at Yon Hall Room 427, 1908 Stadium Road, (352) 273-1094, title-ix@ufl.edu

Sexual Assault Recovery Services (SARS)

Student Health Care Center, 392-1161.

University Police Department at 392-1111 (or 9-1-1 for emergencies), or http://www.police.ufl.edu/.

COVID-19

- You are expected to wear approved face coverings at all times during class and within buildings even if you are vaccinated.
- If you are sick, stay home and self-quarantine. Please visit the UF Health Screen, Test & Protect website about next steps, retake the questionnaire and schedule your test for no sooner than 24 hours after your symptoms began. Please call your primary care provider if you are ill and need immediate care or the UF Student Health Care Center at 352-392-1161 (or email covid@shcc.ufl.edu) to be evaluated for testing and to receive further instructions about returning to campus.
- If you are withheld from campus by the Department of Health through Screen, Test & Protect, you are not permitted to use any on campus facilities. Students attempting to attend campus activities when withheld from campus will be referred to the Dean of Students Office.
- UF Health Screen, Test & Protect offers guidance when you are sick, have been exposed to someone who has tested positive or have tested positive yourself. Visit the UF Health Screen, Test & Protect website for more information.

- Please continue to follow healthy habits, including best practices like frequent hand washing. Following these practices is our responsibility as Gators.

Academic Resources

E-learning technical suppor*t*, 352-392-4357 (select option 2) or e-mail to Learning-support@ufl.edu. https://lss.at.ufl.edu/help.shtml.

Career Resource Center, Reitz Union, 392-1601. Career assistance and counseling; https://career.ufl.edu.

Library Support, http://cms.uflib.ufl.edu/ask. Various ways to receive assistance with respect to using the libraries or finding resources.

Teaching Center, Broward Hall, 392-2010 or 392-6420. General study skills and tutoring. https://teachingcenter.ufl.edu/.

Writing Studio, 302 Tigert Hall*,* 846-1138. Help brainstorming, formatting, and writing papers. https://writing.ufl.edu/writing-studio/.

Student Complaints Campus*:* https://sccr.dso.ufl.edu/policies/student-honor-code-student-conduct-code/;https://care.dso.ufl.edu.

On-Line Students Complaints*:* http://www.distance.ufl.edu/student-complaint-process.