Exploring Data in R & Python

DAT 301

Instructor Info

Shiwei Lan

Office Hrs: TTh 10:30-11:30 AM

WXLR 544

https://math.asu.edu/~slan

slan@asu.edu

Course Info

Prereq: MAT 266 (Cal II) and MAT 343 (linear algebra)

TTh 9:00 – 10:15 AM

Tempe LIB C5

https://slan-teaching.github.io/DAT301/

Grader Info

Shuyi Li

Office Hrs: TBA

Virtual, TBA

shuyili3@asu.edu

Discription

This course focuses on first steps in data processing and exploratory data analysis, which include: data cleaning (tidying), data manipulation (wrangling, munging), exploratory data analysis (visualization, plots, summaries), as well as some statistical procedures for model fitting and predicting (linear regression models, trees, clustering).

Objectives

By the end of the course students should be able to perform basic data cleaning and manipulation using `base R` and `dplyr`; use `base R`, `ggplot2` and `plotly` packages to create plots; be able to create dynamic reports and slides in Rmarkdown, PDF and HTML formats using `knitr` (see example here); create `shiny` interactive apps see example here; be able to make simple animations using packages `animation` and `gganimate` (such as the one below, from Rosling's `gapminder` data). In Python, students should be able to use Jupyter notebook for coding and HTML/PDF reports; use libraries `numpy`, `pandas`, `matplotlib` and `seaborn` for data wrangling and visualisation; use `BeautifulSoup` for a simple webscraping.

Textbooks

Required

R4DS - R for Data Science (free online https://r4ds.had.co.nz) by Hadley Wickham and Garrett Grolemund

Python4DS - Python for Data Analysis (2nd edition) by Wes McKinney

Recommended

R&Py - Exploring Data in R and Python (online ebook) by Marko Šamara

Grading Scheme

Homework $\times 5$	25%
Labs $\times 5$	25%
Project 1 (midterm)	25%
Project 2 (final)	25%
Total	100%

A+	[97%, 100%]	A	[93%, 97%)	A-	[90%, 93%)
B+	[87%, 90%)	B	[83%, 87%)	B-	[80%, 83%)
C+	[77%, 80%)	C	[70%, 77%)		
D	[60%, 70%)			E	[0%, 60%)

Homework

There will be ≈ 5 HW assignments, done in Jupyter notebook. They test your coding (both in `R` and `Python`). A software automatically grades homework (so, no partial credit within a code cell!). You will have about a week for each homework, so please plan accordingly. Leaving the town for couple of days (such as for going to a wedding, funeral, conference, court, etc.) is NOT an excuse for not finishing the homework on time.

Labs

There will be ≈ 5 labs, which consist of programming in `R` and `Python`. While homework assignments are more instructed with exact expectation from a student, labs are intended to be mini projects, with less strict instructions, and more freedom for students to decide what tools to use and what to have as an outcome. Just as HW, no late lab will be accepted.

Projects

Project 1 serves as a midterm, while Project 2 as the final exam (there is no in-class midterm or final exam). Students are encouraged to work in groups (up to 3) on projects. Each group would submit code, the outcome of the code, and for Project 1, presentation slides as well. The slides will be shown and the analysis presented to the class. (It should be fun ☺). Project 2 has similar format as Project 1, except that there will be no presentation. Project 2 must be submitted by 12/10/2020.

FAQs

Where can I find help?

You can go to my virtual office hours and the grader's office hours. In addition, you can go to slack DAT 301 channel to post your questions and help others.

How do I keep track of the class?

Constantly check canvas and the course website. I will make announcements, post homework solutions, etc..

Do we have incentives?

I will give bonus points through the semester for e.g. extra-credit homework problems, most helpful piazza users, etc..

When shall I drop if I choose to?

Last Day to Register or Drop/Add Without College Approval is 08/26/2020. Tuition & Fees Refund Deadline is 09/02/2020 for session C. Course Withdrawal Deadline (without 'W' on your transcript) is 11/04/2020 for session C. Refer to https://students.asu.edu/academic-calendar for more deadlines.

Software

We will be using programming languages R and Python. For R, we will use RStudio as an IDE, while for Python we will use Jupyter and Spyder. You can install them by installing Anaconda for free (see instructions in the course material). You can also use computers in ASU's libraries for working on the assignments.

- SoMSS computer labs. The computers run Linux systems and have R installed. Check the following website for open hours and more information: https://math.asu.edu/resources/computing-resources
- Computing Commons and GWC lab. Check https://uto.asu.edu/computing/tempe for more info.

Disability Accommodations

Qualified students with disabilities are encouraged to make their requests at the beginning of the semester to get disability accommodations. Disability information is confidential. Note: *Prior to receiving disability accommodations, verification of eligibility from the Disability Resource Center (DRC) is required.* Therefore, you should contact DRC immediately. Their office is located on the first floor of the Matthews Center Building. DRC staff can also be reached at: 480-965-1234 (V), 480-965-9000 (TTY). For additional information, visit: www.asu.edu/studentaffairs/ed/drc. Their hours are 8:00 AM to 5:00 PM, Monday through Friday.

Cell phones and Electronic Devices

Picture taking, talking or texting on your cell phone or any electronic device during class is prohibited. If you bring a cell phone and/or any other electronic equipment to the class, make sure they are turned off before class begins. Any sounds produced by such devices are disruptive to the class and, as such, will not be tolerated and may be reported to the Office of the Dean of Students.

Academic Honesty

ASU expects and requires all its students to act with honesty and integrity, and respect the rights of others in carrying out all academic assignments. For more information on academic integrity, including the policy and appeal procedures, please visit http://provost.asu.edu/academicintegrity.

Inclusion

The School of Mathematical and Statistical Sciences encourages faculty to address and refer to students by their preferred name and gender pronoun. If your preferred name is different than what appears on the class roster, or you would like to be addressed using a specific pronoun, please let me know.

Sexual Violence and Harassment

Both Title IX federal law and university policy make clear that sexual violence and harassment based on sex is prohibited. An individual who believes they have been subjected to sexual violence or harassed on the basis of sex can seek support, including counseling and academic support, from the university. If you or someone you know has been harassed on the basis of sex or sexually assaulted, you can find information and resources at https://sexualviolenceprevention.asu.edu/faqs. As a mandated reporter, I am obligated to report any information I become aware of regarding alleged acts of sexual discrimination, including sexual violence and dating violence. ASU Counseling Services, https://eoss.asu.edu/counseling, is available if you wish to discuss any concerns confidentially and privately.

Syllabus Disclaimer

This syllabus is tentative and should not be considered definitive. The instructor reserves the right to modify it (including the dates of the tests) to meet the needs of the class. Every effort will be made to avoid changing the course schedule but the possibility exists that unforeseen events will make syllabus changes necessary. It is the student responsibility to attend class regularly and make note of any change.

FAQs

Where can I find help?

You can go to my virtual office hours and the grader's office hours. In addition, you can go to slack DAT 301 channel to post your questions and help others.

How do I keep track of the class?

Constantly check canvas and the course website. I will make announcements, post homework solutions, etc..

Do we have incentives?

I will give bonus points through the semester for e.g. extra-credit homework problems, most helpful piazza users, etc..

When shall I drop if I choose to?

Last Day to Register or Drop/Add Without College Approval is 08/26/2020. Tuition & Fees Refund Deadline is 09/02/2020 for session C. Course Withdrawal Deadline (without 'W' on your transcript) is 11/04/2020 for session C. Refer to https://students.asu.edu/academic-calendar for more deadlines.

Class Rules in response to COVID-19

- Immersion modality: This class will follow the Immersion model of delivery. Classes will be held in person in our scheduled room and at the scheduled time, with a simultaneous remote delivery via Zoom. Your instructor will specify how the in-person arrangements will occur. If you are unable to attend in person, please notify your instructor (from your ASU email account) in a timely manner to make them aware.
- Zoom Etiquette: During the Zoom sessions, please log in on time and assure that you have a reasonably secure connection. Please use your full name or first name-last initial. No outside attendees will be allowed, and during the sessions, please keep your microphone's audio muted except when needing to talk to the instructor. The instructor reserves the right to remove anyone from the Zoom sessions for disruptive behavior.
- Masks: **IMPORTANT!** For sessions held in person, masks and other personal protection equipment (PPE) **MUST BE WORN** in accordance with ASU's policies. Failure to do so will result in your being asked to comply, then asked to leave if unable or unwilling to comply. Deliberate refusal to comply will be treated as a Student Code of Conduct violation and referred to the Dean's office for review.

Class Schedule

Week	Date	Topic	Assignments
1	08/20 - 08/21	Introduction	
2	08/24 - 08/28	Introduction to R	HW 1 out
3	08/31 - 09/04	Random Variables and Data / Basic Plots	HW 1 due; LAB 1 out
4	09/08 - 09/11	Basic Workflow/ Apply Family	LAB 1 due
5	09/14 - 09/18	Data Manipulation / RMarkdown	HW 2 out
6	09/21 - 09/25	Regression	HW 2 due; LAB 2 out
7	09/28 - 10/02	Ilslides and Plotly	LAB 2 due
8	10/05 - 10/09	ggplot2 / animation	HW 3 out; LAB 3 out
9	10/12 - 10/16	Project Presentations	project 1 (midterm): due 10/18/2019
10	10/19 - 10/23	Shiny	HW 3 due; LAB 3 due
11	10/26 - 10/30	iPython Notebook	HW 4 out
12	11/02 - 11/06	Data Structures and Function	HW 4 due; LAB 4 out
13	11/10 - 11/13	Numpy and Matplotlib	LAB 4 due
14	11/16 - 11/20	Pandas	HW 5 out
15	11/23 - 11/25	Web Scraping	HW 5 due; LAB 5 out
16	11/30 - 12/04	Matplotlib and Seaborn	LAB 5 due
Final	12/07 - 12/12	Final Exam	project 2 (final): due 12/10/2020