



Exploring Data in R & Python

DAT 301

Instructor Info —



Shiwei Lan



Office Hrs: MW 12:00-1:00 PM



WXL R 544; Zoom: 8055899886



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



slan@asu.edu

Course Info —



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



MW 1:30 – 2:45 PM



WXL R A108; Zoom pwd 94171



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

Grader Info —



Alejandro Vidales Aller



Office Hrs: TBA



Virtual:

<http://asu.zoom.us/my/mathtutoring>



avidales@asu.edu

Description

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 ganimate (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

Written HW ×5	25%
Coding HW ×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%]

Written Homework

There will be ≈ 5 written assignments on canvas, which consist of programming in R and Python. Each homework report should be submitted in either Word or PDF format, no other formats accepted. Late home will NOT be accepted. Do NOT send your homework by email!

Coding Homework

There will be ≈ 5 coding assignments on NBGrader, a platform based on Jupyter notebook that automatically grades homework. They test your coding (both in R and Python). You will have about a week for each homework, so please plan accordingly. Late submission will NOT be graded.

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 08/24/2022. Tuition & Fees Refund Deadline is 08/31/2022 for session C. Course Withdrawal Deadline (without 'W' on your transcript) is 11/02/2022 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.

Class Schedule

Week	Date	Topic	Assignments
1	08/18 -	Instruction begins	
2	08/22 - 08/26	Introduction to R	Coding 1 out
3	08/29 - 09/02	Random Variables and Data / Basic Plots	Coding 1 due; Written 1 out
4	09/06 - 09/09	Basic Workflow/ Apply Family	Written 1 due
5	09/12 - 09/16	Data Manipulation / RMarkdown	Coding 2 out
6	09/19 - 09/23	Regression	Coding 2 due; Written 2 out
7	09/26 - 09/30	Ilslides and Plotly	Written 2 due
8	10/03 - 10/07	ggplot2 / animation	Coding 3 out; Written 3 out
9	10/10 - 10/14	Project Presentations	project 1 (midterm): due 10/16/2019
10	10/17 - 10/21	Shiny	Coding 3 due; Written 3 due
11	10/24 - 10/28	iPython Notebook	Coding 4 out
12	10/31 - 11/04	Data Structures and Function	Coding 4 due; Written 4 out
13	11/08 - 11/11	Numpy and Matplotlib	Written 4 due
14	11/14 - 11/18	Pandas	Coding 5 out
15	11/21 - 11/23	Web Scraping	Coding 5 due; Written 5 out
16	11/28 - 12/02	Matplotlib and Seaborn	Written 5 due
Final	12/05 - 12/10	Final Exam	project 2 (final): due 12/10/2020