HSS 611: Programming for Digital Humanities and Computational Social Sciences

Fall 2024
Mon Wed 2:30–4:00pm
N4 1309, School of Digital Humanities and Computational Social Sciences

Instructor: Taegyoon Kim, Ph.D. in Political Science and Social Data Analytics

- Email: taegyoon@kaist.ac.kr
- Office hours: Tue Thu 11:00–12:00pm & By appointment
- Personal webpage: https://taegyoon-kim.github.io

Course Overview: This course offers a comprehensive introduction to programming tailored for data analytics in humanities and social sciences. Students will learn the core principles of programming along with essential skills for quantitative and computational research. Key topics include—but are not limited to—Python fundamentals, data manipulation, web data collection, text-as-data/NLP, network analysis, and machine learning. The course integrates theoretical discussions with hands-on coding exercises, providing students with practical experience. Additionally, through applied projects, students will have the opportunity to apply and adapt their newly acquired skills to their own research.

Readings: Students are not required to purchase any books for this course as it will not follow any textbook strictly. However, here are useful, freely available books:

- McKinney, W., 2022. Python for data analysis. "O'Reilly Media, Inc." [link]
- VanderPlas, Jake. 2016. Python data science handbook: Essential tools for working with data. "O'Reilly Media, Inc." [link]
- Wickham, H., Çetinkaya-Rundel, M. and Grolemund, G., 2023. R for data science. "O'Reilly Media, Inc.". [link]

Major Tasks: The course materials (lecture slides, exercises and example solutions, etc.) will be uploaded to the course Github repository. Students are expected to complete the following tasks.

- *Attendance:* Students are required to attend all lectures unless they have exceptional circumstances that they have discussed with the instructor beforehand. Worth 10% of the final grade.

- *Exercises:* There will be several exercises for Python fundamentals. The due dates are indicated in the weekly schedule below. Upload your completed exercise to the `exercises` folder. Worth 20% of the final grade.

- *Tutorial presentation:* Students will deliver one tutorial presentation for approximately 10–15 minutes. Worth 20% of the final grade.

    - The presentations serve a dual purpose: firstly, to provide non-presenting students with a deeper understanding of the techniques covered in that week, and secondly, to offer presenting students an opportunity to apply and adapt those techniques in their own data analysis tasks.
    - At the beginning of the tutorial, provide a clear overview that includes what the tutorial covers, the specific data being used, and any other relevant details.
    - During the presentations, students will showcase the implementation of techniques covered in the previous week using their own data, preferably related to their research projects.
    - They will guide the entire class through their script, explaining each line of code in detail, including the syntax when necessary.
    - Tutorial presentations will be on Wednesday classes. Please sign up on this spread sheet.
    - Students are free to use any scripting tool, but Google Colaboratory is an effective choice for creating and presenting tutorials.
    - Be sure to provide the class with access to your script and data so that the whole class can follow as you walk through your tutorial by uploading them to the `tutorial_materials` folder prior to the presentation.

- *Final project:* Students must complete an original research paper that involves a significant amount of data work and present it during the final week of the course. The paper and presentation are worth 40% and 10% of the final grade, respectively.

    - The primary objective is to guide students in building a comprehensive data-driven research pipeline that encompasses data collection, preprocessing, analysis, interpretation, and visualization.

- Although students have the freedom to choose their preferred topic, they are required to work with data applying skills learned in the course. By applying and adapting these skills for their own research, students will not only demonstrate their mastery of programming fundamentals and data analytic skills but also gain a practical understanding of how what they learned in class can be effectively utilized in real-world research scenarios.
- The project should focus on detailed methodological accounts of data collection, cleaning, manipulation, analysis, and visualization, rather than discussions of theories or prior literature.
- Students will deliver a short presentation of their papers (10 minutes) in the final week of the course. The paper should be around 3,000 words in length—excluding figures, tables, and appendices—and should be uploaded, along with replication materials, to the the `final_projects` folder by the end of the semester (Dec 20 11:59pm).

Grading Scale: Grade values will not be rounded. That is, any grade value that is greater than or equal to 'Lower' and less than 'Upper' will receive the respective grade.

Grade	Lower	Upper
A+	90	101
A_0	87	90
A-	84	87
B+	81	84
B_0	78	81
B-	75	78
C+	72	75
C_0	69	72
C-	66	69
D+	63	66
D_0	60	63
D-	57	60
F	0	57

Course Schedule:

Week 1. Sep 2 & 4

- Mon: Course overview, logistics, self-introductions, etc.
- Wed [Zoom [link]]: Branching and iteration
  - ⋆ Exercise 1 (due date: 11:59pm Sep 11)

Week 2. Sep 9 & 11

- Mon: No class (instructor at conference)
- Wed: String manipulation, tuples, and lists
  - ⋆ Exercise 2 (due date: 11:59pm Sep 23)

Week 3. Sep 16 & 18

- Mon & Wed: No class (Chuseok)

Week 4. Sep 23 & 25

- Mon: Functions, modules, and exceptions
  - ⋆ Exercise 3 (due date: 11:59pm Sep 30)
- Wed: Dictionaries, sets, and recursion
  - ⋆ Exercise 4 (due date: 11:59pm Oct 2)

Week 5. Sep 30 & Oct 2

- Mon: NumPy and Pandas I
- Wed: NumPy and Pandas II
  - ⋆ Exercise 5 (due date: 11:59pm Oct 14)

Week 6. Oct 7 & Oct 9

- Mon: TBD
- Wed: No class (Hangul Day)

Week 7. Oct 14 & Oct 16

- Mon: One-on-one proposal meetings I

- Wed: One-on-one proposal meetings II

Week 8. Oct 21 & 23

- Mon & Wed: No class (mid-term examination period)

Week 9. Oct 28 & 30

- Mon & Wed: Visualization

Week 10. Nov 4 & Nov 6

- Mon & Wed: Interacting with API (Application Programming Interface)

Week 11. Nov 11 & 13

- Mon & Wed: Web scraping

Week 12. Nov 18 & 20

- Mon & Wed: Text-as-data/NLP I

Week 13. Nov 25 & 27

- Mon & Wed: Text-as-data/NLP II

Week 14. Dec 2 & 4

- Mon & Wed: Network analysis

Week 15. Dec 9 & 11

- Mon & Wed: Machine learning

Week 16. Final presentation (Dec 16 & 18)

- Mon & Wed: Final presentations

Instruction Mode: The instruction mode is in-person. However, some classes might be offered remotely. Any change to the mode of instruction will be announced in advance.

Attendance: Consistent attendance is essential for this course. You are permitted to miss a maximum of one class without any impact on your grade. However, should you exceed

this limit and miss additional classes, a deduction of two points will be applied to your attendance score for each absence. Arriving 5 to 20 minutes late will be marked as late. Please note that arriving more than 20 minutes late will be counted as an absence for that class.

Email Policy: I try to respond to emails promptly, typically within two business days. If you have complex questions or need an in-depth discussion, I encourage you to attend my office hours.

Office Hours: I welcome all students to attend my office hours for discussions related to course content and learning strategies. If you need to set up a meeting outside my office hours, send me an email with your availability, and we will arrange a mutually convenient time to meet.

Late Submission Policy: Late submissions will incur a penalty of 10% for each day (rounded up) beyond the due date

Academic Integrity: As students at KAIST, you are entrusted with upholding the utmost standards of academic integrity. Academic honesty is paramount, and any form of misconduct is strictly prohibited. In the event of suspected misconduct, our class adheres to the established policy of KAIST. All such incidents are promptly reported to the dean of the Department of Humanities and Social Sciences to ensure a fair and transparent resolution.

Grade Dispute Memo: Should any student want to contest a grade received on the response note, theme presentation, and team project, they have the option to do so by submitting a written memo. The memo must specifically outline the reasons why the assignment warrants a different grade. The memo must then be submitted within seven days of receiving the grade. Submissions should be sent via email for consideration.

Syllabus Change Policy: This syllabus is a guide, and every attempt will be made to provide an accurate overview of the course. However, circumstances and events may make it necessary for the instructor to modify the syllabus during the semester and may depend, in part, on the progress, needs, and experiences of the students.