API 231
Geographic Information Systems for Public Policy

Harvard Kennedy School | Spring 2024

WHERE & WHEN

LITTAUER 230 (HKS)
TuTh 1030-1145

WHO

YURI M. ZHUKOV
Visiting Assoc Prof of Public Policy
Harvard Kennedy School (2023-24)

Associate Professor
School of Foreign Service
Department of Government
Georgetown University

Contact
email: yzhukov@hks.harvard.edu
website: zhukovyuri.github.io

Office Hours
Tu 1500-1700 (via Zoom)
Signup sheet (24 hours in advance):
calendly.com/zhukov-hks

QUICK LINKS

Course Canvas page
canvas.harvard.edu/courses/132841

Readings
... /files/folder/Readings
Lecture Slides
... /files/folder/Slides
Assignments
... /assignments

Note: No textbook purchases or journal subscriptions/licenses needed. All assigned readings (incl. recommended readings) will be available to enrolled students in PDF format via the course Canvas page.

WHAT

COURSE DESCRIPTION

This course provides and introduction to geospatial data science and its applications to public policy, political economy, and international relations.

The course will enable students to work with Geographic Information Systems (GIS) data structures and understand cartography, transformations, georeferencing, geocoding, and spatial analysis. It will provide students with hands-on experience in using open-source GIS software to visualize and analyze multiple types of data, including administrative records, historical maps, boundaries, weather, climate, roads, and satellite images/remote sensing.

Applications will include data on elections, armed conflict, crime, public infrastructure, and economic activity, plus a series of specialized policy topics selected by students.

Class meeting time will be split between lectures and lab exercises, where students will use GIS software to complete assignments.

GOALS

The learning objectives are twofold. The first is to equip students with the methodological tools needed to collect, pre-process and analyze geospatial data. The second is to provide students with hands-on research experience, in which they will apply their newly-acquired methodological skills to answer a political, social, or economic policy question.

The desired "end state" for this class is for students to become sufficiently proficient in geospatial data science to continue applying and learning these methods on their own. In other words, we can't transform you into a GIS expert overnight. But we will give you the tools you need to become one.

PREREQUISITES

This course has no formal prerequisites.

Prior coursework in statistical computing and exposure to QGIS, R and/or Python may be helpful in digesting some of the topics. However, students are not expected to have quantitative skills or programming language proficiency to enroll in the course.

FORMAT

The course is organized into two 75-minute meetings per week. Meetings will feature a combination of lectures, discussion and activity-based learning through participation in computational exercises and tutorials.

The first part of the class will be a "methods boot camp," with weekly lectures and problem sets on key geospatial data science topics.

COURSE POLICIES

ACADEMIC INTEGRITY

Students are expected to:

- ✓ Follow HKS Academic Code as specified in Student Handbook (tinyurl.com/hkshandbook).
- ✓ Properly cite books, articles, websites, lectures, datasets, and other sources in written work.

  This is especially important for:

  - Text taken, in whole or in part, from writing/speech published elsewhere.
  - Arguments/insights drawn from writing/speech published elsewhere.
  - Code borrowed from online tutorials, textbooks, or replication archives.

Students are encouraged to:

- ✓ Collaborate on a co-authored final project or paper.
- ✓ Consult with classmates on choice of project topics and share sources.
- ✓ Use the final project as an opportunity to further develop ideas for an existing paper or thesis (w/ instructor approval).

The second part of the class will focus on applying this methodological toolkit to hands-on research and analysis projects. Instead of weekly problem sets, students will begin work on their own research projects. Lab exercises will be "walk-throughs" of data collection and analysis on a series of specialized topics, selected by students.

Data analysis walk-through topics (students will vote on 3 out of 10):

1. Agriculture and crop productivity.
2. Congressional redistricting.
3. Climate-conflict nexus.
4. Crime and policing.
5. International migration.
6. Nighttime luminosity.
7. Piracy and transnational shipping.
8. Political repression.
9. Racial and ethnic segregation.
10. Russian-Ukrainian War.

In each walk-through, the professor will provide a "farm-to-table" tour on: (a) where to find and download data, (b) how to pre-process and integrate the data to make them analysis-ready, and (c) how to conduct a very rudimentary analysis, focused on description, explanation or prediction.

SOFTWARE

We will use QGIS and/or R for all lab exercises.

QGIS is a free, open-source, cross-platform software environment for the processing and analysis of geospatial data. It has functionality similar to ESRI's ArcGIS, but without the licensing fees. A background in GIS is helpful, but not required. Students who would like to get a head start are encouraged to download the software here (qgis.org), and consult the introductory tutorial (docs.qgis.org/3.34/en/docs/training_manual/).

Students will also have the option of using the R statistical programming language for most tutorials. R is a free, cross-platform software environment for statistical computing and graphics. Students who would like to get a head start are encouraged to download the software here (http://cran.us.r-project.org/), and consult the introductory tutorial (http://cran.r-project.org/doc/manuals/R-intro.pdf). Some students may prefer the slightly more user-friendly GUI, R Studio (http://www.rstudio.com/). Code and data for all tutorials will be made available through the course website.

GRADES

Grades will be based on problem sets (40%), a final project (40%), and class participation (20%).

1. **Problem sets** (40%). There will be 8 weekly problem sets (5% each), in which students will learn to implement basic GIS data management and analysis methods. Each problem set will follow an in-lab demonstration of key tasks, and an overview of the assignment. Students are encouraged to work collaboratively on these problem sets. The assignments will be due no later than 11:59 PM on the Sunday following each lab (e.g. Sunday Feb. 4 for the Feb. 1 lab, etc.).

STUDENT EVALUATIONS

Student feedback is critical to ensuring that course content is accessible and that we are meeting our learning goals. To this end, API-231 will supplement standard, Harvard-administered end-of-semester evaluations with **lecture-specific** and **lab-specific student evaluations**.

These daily surveys are 100% anonymous and 100% optional. We will use them to make mid-course adjustments and to fine-tune future course materials to better meet student needs.

A link (QR code) to each survey will be posted on Canvas, in the discussion entry for each lecture and lab.

2. **Final project** (40%). Each student will work on an individual research project, in which they will use GIS software and data to answer a political, social or economic question. This question may be descriptive, and oriented toward map-making and visualization (e.g. "Which counties flipped from Democratic to Republican in 2022?", "Which neighborhoods are the most violent?"). The question may also be explanatory, and oriented toward the analysis of geospatial data (e.g. "Why did some counties flip from Republican to Democrat?", "Why are some neighborhoods more violent than others?").

   Each student will submit a 1-paragraph project abstract by 11:59 PM, Friday, March 8 (worth 5% of grade), summarizing the research idea and the spatial and non-spatial data needed for its execution.

   Each student will give a 5-minute class presentation (5% of grade) on April 23 and 25 (last two days of class). This presentation does not require that the project be complete, but must contain 2 slides:

   a) *Research Question*: what question are you asking, and why is it interesting/important?
   b) *Maps*: map(s) indicating the sort of GIS data and methodology you will be using

   Students will also submit a 5-7 page final report (excluding maps, figures and tables). These reports will be due (via Canvas) by 11:59 PM on Friday, May 3 (worth 30% of grade). The report must contain three sections:

   a) *Research Question*: what question are you asking, and why is it interesting/important?
   b) *Data*: what kind of geospatial data did you use to answer this question, where did you find these data, what processing steps were needed to import/transform/merge the data?
   c) *Preliminary results*: what did you find? (map, table or statistical graphic, with a 2-3 page discussion/summary)

3. **Classroom attendance and participation** (20%). Students are expected to engage the readings fully and actively participate in all class discussions.

USE OF GENERATIVE AI FOR COURSEWORK

Generative Artificial Intelligence (AI) is a powerful new tool for learning and discovery. It can be difficult to fully grasp AI's potential impact on students' future careers, but this technology clearly has the potential to increase productivity in certain domains. As a pedagogical tool for developing students' critical thinking and skills, however, AI cannot substitute the value of working through course assignments and readings on one's own.

This course will follow HKS's general policy on students' use of generative AI for coursework: it's OK to use this technology as a tool to look up information, but not OK to use AI to complete assignments. Specifically,

1. It is generally acceptable to use generative AI for finding information and for solidifying knowledge of the course content. For instance, it is fine to use AI-powered web search and to have "conversations" with tools like ChatGPT to help explore ideas, refine one's thinking, identify examples, and better understand the course material. However, unless otherwise specified, it is a violation of the HKS Academic Code to incorporate into one's coursework text produced predominantly by generative AI — similar to copying from a book or article. For example, it would be inappropriate to simply rewrite a draft generated by AI.

2. Generative AI can produce false or misleading information. Students are ultimately responsible for the accuracy of any work they submit. Any facts that students look up using AI chatbots (as with any other online source) should be verified and properly cited with a second, authoritative source (e.g. primary document, news story, academic article or book, administrative records, reference volume). As a general rule of thumb, when looking up information, AI is a good place to start, but the wrong place to end (much like Wikipedia).
3. If students have any doubt about whether a specific use of generative AI is permitted for an assignment or course, they should discuss it with the course instructor prior to using it.

IMPORTANT DATES

SHOPPING DAY
Thursday, Jan 18

COURSE BEGINS
Tuesday, Jan 22

COURSE ADD DEADLINE
Friday, Feb 2

COURSE DROP DEADLINE
(WITHOUT NOTATION)
Monday, Feb 5

CLASS SCHEDULE

INTRODUCTION

0. GIS 101 . . . . . Tuesday, Jan 22
    - Course introduction
    - Overview of GIS software and data structures

    Readings:
    - Ballas, Dimitris, Graham Clarke, Rachel S. Franklin, and Andy Newing. *GIS and the social sciences: Theory and applications*. Routledge, 2017: Ch. 1.

1. *Lab exercise* / problem set 1 . . . . . Thursday, Jan 24
    - Make your first map (elections).

PART I: LEARNING METHODS

2. Map projections and overlays . . . . . Tuesday, Jan 30
    - Coordinate systems and projections.

    Readings:
    - Campbell, Jonathan and Michael Shin. *Essentials of Geographic Information Systems*. Saylor Foundation, 2011: pp. 41-50.
    - Bolstad, Paul. *GIS Fundamentals: A First Text on Geographic Information Systems*. 6th ed. XanEdu Publishing, 2019: Ch. 3. [recommended]

    - Mapping basics.

    Readings:
    - Ballas, et al. (2017): Ch. 3.

3. *Lab exercise* / problem set 2 . . . . . Thursday, Feb 1
    - Plot multiple datasets on same map (violence in Afghanistan).
4. Spatial analysis and geoprocessing . . . . . Tuesday, Feb 6
    - Spatial queries.

    Readings:
    - Ballas, et al. (2017): Ch. 2.
    - Campbell and Shin (2011): Ch. 6-7.

5. *Lab exercise* / problem set 3 . . . . . Thursday, Feb 8
    - Test a hypothesis with spatial data (NYC cycling safety).
6. Turning maps into data (part 1) . . . . . Tuesday, Feb 13
    - Geo-referencing and vectorization.

      Readings:
        - Affek, Andrzej. "Georeferencing of historical maps using GIS, as exemplified by the Austrian Military Surveys of Galicia." *Geographia Polonica* 86, no. 4 (2013): 375-390.
        - Rumsey, David, and Meredith Williams. *Historical maps in GIS*. David Rumsey Historical Map Collection, 2002: Ch. 1.
7. No class meeting today . . . . . Thursday, Feb 15
8. *Lab exercise* / problem set 4 . . . . . Tuesday, Feb 20
    - Geo-reference a map you found online (World War I front line).
9. *Lab exercise* / problem set 5 . . . . . Thursday, Feb 22
    - Create and edit polygon features (historical borders).
10. Turning data into maps . . . . . Tuesday, Feb 27
    - Geo-coding.

      Readings:
        - Goldberg, Daniel W., John P. Wilson, and Craig A. Knoblock. "From text to geographic coordinates: the current state of geocoding." *URISA* 19, no. 1 (2007): 33-46.
11. *Lab exercise* / problem set 6 . . . . . Thursday, Feb 29
    - Geocode a list of locations (historical lynchings in U.S.).
12. Spatial transformations (part 1) . . . . . Tuesday, Mar 5
    - Changes of geographic support.

      Readings:
        - Zhukov, Yuri M., Jason S. Byers, Marty Davidson, and Ken Kollman, "Integrating Data Across Misaligned Spatial Units," *Political Analysis* 32, no. 1 (2024): 17-33.

FINAL PROJECT
Topic selections due
Friday, Mar 8

13. *Lab exercise* / problem set 7 . . . . . Thursday, Mar 7
    - Integrate incompatible vector datasets (elections).

SPRING BREAK
Mar 9 – 17

14. Spatial transformations (part 2) . . . . . Tuesday, Mar 19
    - Remote sensing and satellite data.

      Readings:
        - Crews, Kelley A., and Stephen J. Walsh. "Remote sensing and the social sciences." *The Sage Handbook of Remote Sensing* (2009): 437-445.
15. *Lab exercise* / problem set 8 . . . . . Thursday, Mar 21
    - Integrate raster and vector datasets (weather and climate).

PART II: APPLYING METHODS

16. **Building a GIS research project from the ground up**. Tuesday, Mar 26
    - Research design.

      Readings:
      - Brinton, Willard. *Graphic presentation.* Brinton Associates, 1939: Ch. 18-29. [skim]

17. *Lab exercise* / data analysis walk-through . . . . . Thursday, Mar 28
    - Zero-to-paper in 60 minutes (coalition airstrikes in Syria & Iraq).

18. **Where to find (free, open-source) geospatial data?** . . . Tuesday, Apr 2
    - Online resources.
    - Library resources.

      Guest speaker:
      - Belle Lipton, GIS Librarian, Harvard Map Collection.

19. No class meeting today . . . . . Thursday, Apr 4
    - Go find data for your project!

20. *Lab exercise* / data analysis walk-through . . . . . Tuesday, Apr 9
    - Topic by student request (see list of choices on p. 2).

21. *Lab exercise* / data analysis walk-through . . . . . Thursday, Apr 11
    - Topic by student request (see list of choices on p. 2).

22. *Lab exercise* / data analysis walk-through . . . . . Tuesday, Apr 16
    - Topic by student request (see list of choices on p. 2).

23. **Troubleshooting session** . . . . . Thursday, Apr 18
    - Professor helps students overcome research roadblocks.

24. **Research fiesta! (part 1)** . . . . . Tuesday, Apr 23
    - Student presentations.

25. **Research fiesta! (part 2)** . . . . . Thursday, Apr 25
    - Student presentations.

COURSE DROP DEADLINE
(WITH NOTATION)
Friday, Apr 19

FINAL PROJECT
STUDENT PRESENTATIONS
Apr 23 – 25

FINAL PROJECT
REPORTS DUE VIA CANVAS
Friday, May 3