March 17, 2024 Version-3.2
Hari Iyer

Zhu Family Cancer Prevention Center Course
Geospatial data science for cancer control and prevention

Description:

Aging societies and changing lifestyles are contributing to unprecedented rises in cancer burden worldwide, making global cancer control a leading public health challenge of our time. Effective global cancer control policy should incorporate knowledge about local contextual factors – features of the places where people live, work and play - that determine who develops cancer and who gets treatment. Decisions regarding where to place oncology services must balance considerations of equity and efficiency. Geospatial data science, a novel methodological field that draws from geography, statistics, and computer science, provides analytic tools that can help us consider the role of these important geographic factors in our work towards reducing the global burden of cancer.

This eight-week online course will introduce students to the essential concepts and analytic tools they need to integrate geospatial data science into their work in global cancer control. Students will learn about different geospatial data types, where to acquire geographic data, and how to analyze these data using open source software and coding platforms. In addition, students will learn from experts at the forefront of this new field, who are applying geospatial data science approaches in their research to understand how geographic features and accessibility influence cancer risk and survival. Whether you are working to eliminate cancer disparities, or to prepare your health care system for disruptions due to climate-related weather events, geospatial data science tools will help advance your goals.

Intended audience:

This course is designed for medical and public health professionals who wish to incorporate geospatial data science into their work. We assume familiarity with basic training in epidemiology and biostatistics. Prior programming experience will be helpful but is not required.

Though the course has a public health focus, we expect students from a wide range of backgrounds will be interested in this topic. Current master's students or recent graduates in public health, economics, or computer/data science may be interested in developing skills in this area to advance their careers. Mid-career policy professionals may also wish to learn about these novel geospatial data sources and analytic approaches. Non-profit and government workers may seek to incorporate geographic data science approaches to better target cancer services to populations in need. By using databases and software that are publicly available, we hope to reach a global audience.

Learning goals:

1. Learn vocabulary and concepts related to geographic data analysis
2. Learn how to search for and acquire geographic data
3. Apply geospatial analytic software to perform the following tasks:
    a. Make maps using geographic data
    b. Link spatial factors to tabular datasets
    c. Test for spatial autocorrelation
4. Learn about applications of geospatial data science for cancer control and prevention
    a. Environmental exposures (e.g. air pollution, chemicals)

b. Neighborhood contextual exposures (e.g. deprivation, segregation, green space, light at night)
c. Geographic accessibility to cancer services (e.g. estimating geographic accessibility in low-resource settings)

Course structure:

This online course will be asynchronous. If students choose, they can just engage with practical exercises and lectures that are most relevant to their work, but we strongly encourage completion of quizzes before advancing between modules. Teaching staff will monitor use and completion of different modules. Feedback and questions from previous iterations of the course have been used to improve content and experience.

Lectures and Readings

The course will consist of recorded lectures to introduce vocabulary and concepts. Lectures will be complimented by recommended readings that provide applications of the concepts and methods discussed in the course.

In addition to lectures, students will have access to structured interviews with leading geospatial data science practitioners who will discuss their research applying geospatial data science approaches to study effects of contextual factors (neighborhood deprivation, built environment, environmental risk factors, health care access) on cancer and health care outcomes.

Practical Exercises/Lab

Weekly practical exercises will be circulated that cover specific geospatial data science topics related to data processing, visualization, and geospatial statistical analysis. Exercises will be described in a handout that can be used as a guide, and through recorded lectures in which an instructor will walk through each step of the handout and show how to perform different tasks. Discussion boards will allow students to request assistance from the teaching team and their peers if they experience difficulties with completing the exercises.

Web Forum Discussions

A moderated web forum is accessible through the Canvas website. We encourage students to post questions about assignments and material. Faculty will review and respond within 48 hours. During the last three weeks, students will be encouraged to read papers on geospatial factors associated with cancer risk, focusing on neighborhood social and physical environments and health care access. Handouts will be provided to students to guide their reading and critical appraisal of these articles, and discussion forums will be available to discuss topics of interest with peers.

Quiz

The quiz should be considered as a self-assessment, students may take quizzes as many times as they wish. Students may consult their notes and course materials to complete the quiz. Short quizzes will be administered for each module. At the end of the course, students who wish to obtain a certificate of completion must complete a 15 question quiz.

Certificates will be available on request following satisfactory completion of lectures, practical exercises, contributions to discussion forums, and quiz.

Required Materials:

All students must have access to a personal or work computer. Students will also be required to download R and R Studio to complete practical exercises. Students must also create accounts to access data from Google Earth Engine.

March 17, 2024 Version-3.2
Hari Iyer

Course Schedule

Week/Topic	Lecture	Exercise	Seminar Speaker (pre-recorded)
1/Course overview	Course introduction • (Lecture 1) Introduce key concepts in geographic analysis (projections, data types, where to access data) • (Lecture 2) Learn where to acquire spatial data	Search for metadata corresponding to geospatial exposure of interest Quiz: Module 1	
2/Data processing with vector data	Overview of R (Lab 1: Vector data processing) • Orientation to platform • Uploading data • Performing tasks	Link census tract socioeconomic features to epidemiologic dataset Optional readings (green space and cancer; geospatial data science resources) Quiz: Module 2	**Peter James ScD** *Associate Professor* *Harvard Medical School &* *Harvard Pilgrim Health Care Institute* **Natural outdoor environments and cancer**
3/Data processing with raster data-1	Overview of Google Earth Engine • Mapping and time series with raster data (Lab 2: Time series/maps) • Focal statistics calculations (Lab 3: Focal statistics)	Data processing using raster data Calculate focal statistics to summarize information from raster data with a given geographic scale Optional readings (geospatial data science resources) Quiz: Module 3	
4/Data processing with raster data-2	Further topics in geospatial data processing (Lab 4) • Raster data: zonal statistics • Raster data: joins, mapping (Lab 5)	Link zonal or focal statistics for NDVI to epidemiologic dataset Quiz: Module 4	**Marcia Castro PhD** *Andelot Professor of Demography* *Harvard T. H. Chan School of Public Health* **Using maps to study malaria transmission in the Amazon**
5/Geospatial statistics	Introduction to geospatial statistical analysis • Conceptual overview (Lecture 3) • Technical exercises (Lab 6)	Evaluate the presence of spatial autocorrelation between two variables (e.g. race and income in	

	○ Spatial autocorrelation ○ Global and local Moran's I	the US, population density and travel time in SSA) Quiz: Module 5	
6/Applications: geographic access	Defining access • Measurement of geographic access • Estimating travel times using Access Mod 5 and Earth Engine • Study design considerations	Optional readings Quiz: Module 6	**Peter Macharia PhD** *Newton International Fellow* *Lancaster University* **Geographic access to health care in low-and middle-income countries**
7/Applications: built and social environment	Defining the built environment • Measurement considerations • Examples of studies of the built environment and cancer • Study design considerations	Optional readings Quiz: Module 7	**Scarlett Lin Gomez PhD** *Professor* *University of California, San Francisco* **Social environments and cancer**
8/Conclusion	Review of course • Why geospatial analysis is important for cancer prevention and control • Tools for geospatial analysis • Applications to improve cancer control and prevention globally	Quiz (for certificate)	

Lecture, Lab, Interview, Quiz

We expect students should spend between 4-5 hours per week to complete lectures and exercises, for a total of 32-40 hours to complete the course.

Readings:

Week 1: Overview & Introduction to Geospatial Data

Reference textbook (R):

Lovelace R, Nowosad, J., & Muenchow, J. (2019). *Geocomputation with R*. Chapman and Hall/CRC Press. Available at: https://geocompr.robinlovelace.net/

Geospatial Data Resources:

The DHS Program - Data [Internet]. [cited 2022 Feb 23]. Available from: https://dhsprogram.com/Data/

Spatial Data Repository - Welcome [Internet]. [cited 2022 Feb 23]. Available from: https://spatialdata.dhsprogram.com/home/

Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment. 2017 Dec 1;202:18–27.

Socioeconomic Data and Applications Center | SEDAC [Internet]. [cited 2022 Feb 23]. Available from: https://sedac.ciesin.columbia.edu/

OpenStreetMap [Internet]. OpenStreetMap. [cited 2022 Feb 23]. Available from: https://www.openstreetmap.org/

Week 2: Data processing with Vector Data in R

Reference textbook (R):

Lovelace R, Nowosad, J., & Muenchow, J. (2019). *Geocomputation with R*. Chapman and Hall/CRC Press. Available at: https://geocompr.robinlovelace.net/

Optional readings:

James P, Banay RF, Hart JE, Laden F. A Review of the Health Benefits of Greenness. Curr Epidemiol Rep. 2015 Jun;2(2):131–42.

James P, Hart JE, Banay RF, Laden F. Exposure to Greenness and Mortality in a Nationwide Prospective Cohort Study of Women. Environ Health Perspect. 2016;124(9):1344–52.

Jimenez MP, DeVille NV, Elliott EG, Schiff JE, Wilt GE, Hart JE, et al. Associations between Nature Exposure and Health: A Review of the Evidence. Int J Environ Res Public Health. 2021 Apr 30;18(9):4790.

James P, Hart JE, Hipp JA, Mitchell JA, Kerr J, Hurvitz PM, et al. GPS-based exposure to greenness and walkability and accelerometry-based physical activity. Cancer Epidemiol Biomarkers Prev. 2017 Apr;26(4):525–32.

Walker, KE. (2023). Analyzing US Census Data: Methods, Maps, and Models in R. CRC Press: Boca Raton, FL. Available at: https://walker-data.com/census-r/index.html#about-the-book

Week 3: Google Earth Engine

Reference resources (Google Earth Engine):

Google Earth Engine. Documentation: JavaScript and Python Guides. Available at: https://developers.google.com/earth-engine/guides

Google Earth Engine. Earth Engine Community Content: Tutorials. Available at: https://developers.google.com/earth-engine/tutorials

Cardille JA, Crowley MA, Saah D, Clinton NE. (2024). Cloud-Based Remote Sensing with Google Earth Engine: Fundamentals and Applications. Springer Cham. Available at: https://link.springer.com/book/10.1007/978-3-031-26588-4

Week 4: Google Earth Engine

Reference resources (zonal and focal statistics):

Fong KC, Iyer HS, James P. Scaling up spatial analyses: Google Earth Engine. Webinar delivered as part of International Society for Environmental Epidemiology North American Chapter Educational Series, 2020. Available at: https://www.youtube.com/watch?v=k7MIy52xIZw

Google Earth Engine. Grouped Reductions and Zonal Statistics. Available at: https://developers.google.com/earth-engine/guides/reducers_grouping

Cardille JA, Crowley MA, Saah D, Clinton NE. (2024). Cloud-Based Remote Sensing with Google Earth Engine: Fundamentals and Applications. Springer Cham. Available at: https://link.springer.com/book/10.1007/978-3-031-26588-4

Optional readings:

de Castro MC, Monte-Mór RL, Sawyer DO, Singer BH. Malaria risk on the Amazon frontier. PNAS. 2006 Feb 14;103(7):2452–7.

Castro MC, Kim S, Barberia L, Ribeiro AF, Gurzenda S, Ribeiro KB, et al. Spatiotemporal pattern of COVID-19 spread in Brazil. Science. 2021 May 21;372(6544):821–6.

Week 5: Spatial Autocorrelation and Statistical Inference

Reference resources (spatial autocorrelation):

Anselin L. Local Indicators of Spatial Association—LISA. Geographical Analysis. 1995;27(2):93–115.

Anselin L. Exploring Spatial Data with GeoDa: A Workbook. Center for Spatially Integrated Social Science: University of Illinois, Urbana-Champaign; 2005. Available at: https://www.geos.ed.ac.uk/~gisteac/fspat/geodaworkbook.pdf

Anselin L. Global Spatial Autocorrelation (1). 2020; GeoDa: An Introduction to Spatial Data Analysis. Available at: https://geodacenter.github.io/workbook/5a_global_auto/lab5a.html

Week 6: Geographic access to care in low-resource settings

Optional readings:

Macharia PM, Mumo E, Okiro EA. Modelling geographical accessibility to urban centres in Kenya in 2019. PLOS ONE. 2021 May 14;16(5):e0251624.

Macharia PM, Odera PA, Snow RW, Noor AM. Spatial models for the rational allocation of routinely distributed bed nets to public health facilities in Western Kenya. Malaria Journal. 2017 Sep 12;16(1):367.

Macharia PM, Ouma PO, Gogo EG, Snow RW, Noor AM. Spatial accessibility to basic public health services in South Sudan. Geospat Health. 2017 May 11;12(1):510.

Alegana VA, Maina J, Ouma PO, Macharia PM, Wright J, Atkinson PM, et al. National and sub-national variation in patterns of febrile case management in sub-Saharan Africa. Nat Commun. 2018 Nov 26;9(1):4994.

Joseph NK, Macharia PM, Ouma PO, Mumo J, Jalang'o R, Wagacha PW, et al. Spatial access inequities and childhood immunisation uptake in Kenya. BMC Public Health. 2020 Sep 15;20(1):1407.

Maina J, Ouma PO, Macharia PM, Alegana VA, Mitto B, Fall IS, et al. A spatial database of health facilities managed by the public health sector in sub Saharan Africa. Scientific Data. 2019 Jul 25;6(1):1–8.

Ouma PO, Maina J, Thuranira PN, Macharia PM, Alegana VA, English M, et al. Access to emergency hospital care provided by the public sector in sub-Saharan Africa in 2015: a geocoded inventory and spatial analysis. The Lancet Global Health. 2018 Mar 1;6(3):e342–50.

Ouma PO, Macharia PM, Okiro EA, Alegana VA. Methods of Measuring Spatial Accessibility to Health Care in Uganda. 2021 Jan 1 [cited 2022 Feb 24]; Available from: https://www.scinapse.io

Iyer HS, Flanigan J, Wolf NG, Schroeder LF, Horton S, Castro MC, et al. Geospatial evaluation of trade-offs between equity in physical access to healthcare and health systems efficiency. BMJ Global Health. 2020 Oct 1;5(10):e003493.

Iyer HS, Wolf NG, Flanigan JS, Castro MC, Schroeder LF, Fleming K, et al. Evaluating urban–rural access to pathology and laboratory medicine services in Tanzania. Health Policy and Planning [Internet]. 2021 Jul 1 [cited 2021 Jul 5];(czab078). Available from: https://doi.org/10.1093/heapol/czab078

Knapp GC, Tansley G, Olasehinde O, Alatise OI, Wuraola F, Olawole MO, et al. Mapping Geospatial Access to Comprehensive Cancer Care in Nigeria. J Glob Oncol [Internet]. 2019 Oct 21 [cited 2020 Sep 28];5. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6825250/

Knapp GC, Tansley G, Olasehinde O, Wuraola F, Adisa A, Arowolo O, et al. Geospatial access predicts cancer stage at presentation and outcomes for patients with breast cancer in southwest Nigeria: A population-based study. Cancer. 2021;127(9):1432–8.

Week 7: The built and social environment and cancer

Optional readings:

Gomez SL, Shariff-Marco S, DeRouen M, Keegan THM, Yen IH, Mujahid M, et al. The impact of neighborhood social and built environment factors across the cancer continuum: Current research, methodological considerations, and future directions. Cancer. 2015 Jul 15;121(14):2314–30.

Gomez SL, Glaser SL, McClure LA, Shema SJ, Kealey M, Keegan THM, et al. The California Neighborhoods Data System: a new resource for examining the impact of neighborhood characteristics on cancer incidence and outcomes in populations. Cancer Causes Control. 2011 Apr;22(4):631–47.

Gomez SL, Clarke CA, Shema SJ, Chang ET, Keegan THM, Glaser SL. Disparities in Breast Cancer Survival Among Asian Women by Ethnicity and Immigrant Status: A Population-Based Study. Am J Public Health. 2010 May 1;100(5):861–9.

Conroy SM, Shariff-Marco S, Yang J, Hertz A, Cockburn M, Shvetsov YB, et al. Characterizing the neighborhood obesogenic environment in the Multiethnic Cohort: a multi-level infrastructure for cancer health disparities research. Cancer Causes Control. 2018;29(1):167–83.

Morey BN, Gee GC, Shariff-Marco S, Yang J, Allen L, Gomez SL. Ethnic enclaves, discrimination, and stress among Asian American women: Differences by nativity and time in the United States. Cultur Divers Ethnic Minor Psychol. 2020 Oct;26(4):460–71.

Iyer HS, Gomez SL, Chen JT, Trinh Q-D, Rebbeck TR. Trends in mortality among Black and White men with prostate cancer in Massachusetts and Pennsylvania: Race and neighborhood socioeconomic position. Cancer. 2021;127(14):2525–34.

Week 8: Course Summary and Wrap-up

Warnecke RB, Oh A, Breen N, Gehlert S, Paskett E, Tucker KL, Lurie N, Rebbeck T, Goodwin J, Flack J, Srinivasan S, Kerner J, Heurtin-Roberts S, Abeles R, Tyson FL, Patmios G, Hiatt RA. Approaching health disparities from a population perspective: the National Institutes of Health Centers for Population Health and Health Disparities. Am J Public Health. 2008 Sep;98(9):1608-15. doi: 10.2105/AJPH.2006.102525. Epub 2008 Jul 16. PMID: 18633099; PMCID: PMC2509592.

Krieger N. Epidemiology and the web of causation: has anyone seen the spider? Soc Sci Med. 1994 Oct;39(7):887-903. doi: 10.1016/0277-9536(94)90202-x. PMID: 7992123.

Alvidrez J, Castille D, Laude-Sharp M, Rosario A, Tabor D. The National Institute on Minority Health and Health Disparities Research Framework. Am J Public Health. 2019;109(S1):S16-S20. doi:10.2105/AJPH.2018.304883