University of Heidelberg

Geospatial Analysis with R for Development (Geo4Dev)
Winter 2023/24

Professor: Richard Bluhm
Email: richard.bluhm@ivr.uni-stuttgart.de
Office hours (Zoom): Tuesdays, 4-5pm, Sign up here!

Teaching assistant: Fernanda Almeida
Email: fernanda.almeida@awi.uni-heidelberg.de

This course provides an overview of how space is used in empirical economics and introduces *R* as a fully self-contained statistical software and Geographic Information System. We will cover the basics of the *R* language, vector and raster data processing in *R*, and more advanced topics. After obtaining the toolkit, we will learn how these techniques are applied in development economics and beyond by replicating a selection of influential papers in economics (e.g., on nighttime lights, slavery in Africa, infrastructure, or conflict).

Course Prerequisite: graduate-level econometrics and statistics; no prior experience with *R* is required but will be helpful.

Learning Objectives
The overall goal of the course is twofold i) to expose students to a literature in development economics that uses geospatial data in innovative ways, and ii) to provide students with the methodological skills needed to assess these papers critically. The participants will learn to "think spatially" and come up with their original research questions involving spatial methods. After completing this course, students should be able to conduct their own spatial analysis with *R* for research or data science jobs. They will develop the following practical skills:

- Handle non-spatial data with *R*
- Write scripts and functions in *R*
- Handle vector and raster data with *R*
- Work with areas, distances, and different projections in *R*
- Understand basic topology and topological relations
- Run advanced geospatial computations in *R*
- Automatically geocode non-spatial data with *R*
- Analyzing and mapping infrastructure
- Calculating cost distances and solve routing problems
- Detect city boundaries using built-up and nighttime lights data
- Work through cases studies on ethnic diversity, slavery, and conflict in Africa
- Learn how space is used for causal identification in economics

Course Logistics

<u>Lectures and Tutorials:</u> Fridays, 11:00 – 3.00 PM (with lunch break), Room 00.010, Campus Bergheim

Books:

- We are not following any textbook, but portions of these three books are helpful as background reading (chapters are indicated in the course outline). They are especially helpful if you had little exposure to *R* before.
- (r-spatial) Lovelace, R., Nowosad, J., and J. Muenchow. 2023. Geocomputation with R. CRC Press. 2nd edition. Available online, for free!
- (r-tidyverse) Wickham, H., Çetinkaya-Rundel, M., and G. Grolemund. 2023. *R* for Data Science. 2nd edition. O'Reilly. Available online, for free!
- (r-programming) H. Wickham. 2019. Advanced *R*. Chapman and Hall/CRC, 2nd edition. Available online, for free!

<u>Software</u>
Everything in this course will be done with *R* and the *RStudio* IDE, freely available for all major platforms. Please install *R* from http://www.r-project.org/ if you are on Windows or Mac or from your package manager if you work on Linux. Please download *RStudio Desktop* (the free version) from https://rstudio.com/products/rstudio/download/ and install both before the first class. Students with zero prior *R* exposure would ideally spend a few minutes playing around with the free chapter from Datacamp's excellent online course (https://www.datacamp.com/courses/free-introduction-to-r).

<u>Discussion Board</u>
The discussion board on Moodle is where everyone can post interesting things (resources on *R* programming, geospatial papers in economics, news with cool maps, etc.) and have conversations outside of class. The board's main purpose for this course is that *you help each other* with questions on coding or understanding certain concepts in a decentralized manner. I will only check occasionally and moderate lightly.

<u>Grading:</u>
- **Assignments:** 40% of the course grade. Issued Fridays after class. Deadline Thursdays 11:59 PM. Best 5 (out of 6) count towards the grade.
- **Replication:** 40% of the course grade. Preferences are due on November 30, 2023. Replication assignments are issued on December 1, 2023. The paper is 10 pages (all in, without cover), about 2500 to 3000 words. Replications (with code) and slides are due on January 28, 2024, 23:59 hrs.
- **Presentations of replications:** 20% of course grade. 15 minutes plus discussion.

<u>Academic Integrity</u>
Students are expected to submit individual and original work. Working in groups on assignments is acceptable, but every student must write up and submit his/her work

(including the original code). Assignments and the take-home exam will be submitted via Moodle. Plagiarism is unacceptable and will result in a failing grade. You are encouraged to use AI tools in moderation and mark their use by providing links to your conversation if your use AI for your report or presentation. Please see my AI policy (on Moodle).

Course Outline

Week 1

<u>Friday 11:00 – 12:30 hrs (October 20): Introduction to course</u>
- Contents
    - Introduction to course
    - Survey of geospatial research in economics
    - Introduction to *R*
- Key readings
    - Donaldson, D. and A. Storeygard. 2016. The View from Above: Applications of Satellite Data in Economics. *Journal of Economic Perspectives*, 30(4), 171–198.
    - Michalopoulos, S. and E. Papaioannou. 2018. Spatial Patterns of Development: A Meso Approach. *Annual Review of Economics*, 10(1), 383–410.
    - Burke, M., Driscoll, A., Lobell, D. B. and S. Ermon. 2021. Using satellite imagery to understand and promote sustainable development. *Science* 371, 371(6535).
- Lab activities
    - First steps in R and Rstudio

<u>Friday 13:30 – 15:00 hrs (October 20): Crash course in R</u>
- Contents
    - Understanding data formats (vectors, matrices, factors, lists, and data frames)
- Key readings
    - Wickham et al. 2023, Chapters 1-4, 26 and 28.
    - (advanced) Wickham 2019, Chapters 2-7
- Lab activities
    - Basic operations in R
    - Assignment 1 issued

Week 2

<u>Friday 11:00 – 12:30 hrs (October 27): Crash course in R</u>
- Contents
    - Understanding objects and functions in R
- Key readings
    - Wickham et al. 2023, Chapters 1-4, 26 and 28.
    - (advanced) Wickham 2019, Chapters 2-7
- Lab activities

- Data wrangling in base R

Friday 13:30 – 15:00 hrs (October 27): Crash course in R (tidyverse)
- • Contents
  - - Introduction to the "tidyverse" package and tidy-paradigm
- • Key readings
  - - Wickham et al. 2023, Chapters 5-9, 13-20, 21.
- • Lab activities
  - - Data wrangling tidy-style
  - - Assignment 2 issued

Week 3

Friday 11:00 – 12:30 hrs (November 3): Introduction to geospatial analysis w/ R
- • Contents
  - - Survey of geospatial data types
  - - Datums, projections, and distances (theory)
  - - Projections and spatial references (in R)
- • Key readings
  - - Lovelace et al. 2023, Chapters 1 and 7.
- • Lab activities
  - - First steps in with sf-package

Friday 13:30 – 15:00 hrs (November 3): Vector data types
- • Contents
  - - Points, lines, polygons, and grids
  - - Reading and writing vector data
  - - Making maps of vector data
  - - Finding and accessing online vector data
- • Key readings
  - - Lovelace et al. 2023, Chapters 2.1–2.2, 2.4, 2.5
- • Lab activities
  - - Calculating road densities and mapping African infrastructure
  - - Projections, areas, and distances
- • Assignment 3 issued

Week 4

Friday 11:00 – 12:30 hrs (November 10): Manipulating vector data
- • Contents
  - - Making points, lines, and polygons
  - - Topology and topological relations
  - - Geometric operations and buffers
  - - Spatial joins and aggregation
- • Key readings
  - - Lovelace et al. 2023, Chapters 3.2, 4.2, 5.2
- • Lab activities
  - - Dealing with invalid polygons

- Road densities at the cell level

<u>Friday 13:30 – 15:00 hrs (November 10): Manipulating vector data (continued)</u>
- Assignment 4 issued

Week 5

<u>Friday 11:00 – 12:30 hrs (November 17): Raster data</u>
- Contents
    - Raster data types and resolutions
    - Reading and writing raster data
    - Cropping, masking, and resampling
    - Extracting, aggregating and disaggregating
    - Finding and accessing online raster data
- Key readings
    - Lovelace et al. 2023, Chapters 2.3, 3.3, 4.3, 5.3, 6
- Lab activities
    - Population and light densities in Kenya
    - Classifying urban areas and finding city footprints

<u>Friday 13:30 – 15:00 hrs (November 17): Raster data (continued)</u>
- Assignment 5 issued

Week 6

<u>Friday 11:00 – 12:30 hrs (November 24): Raster data (continued)</u>

<u>Friday 13:30 – 15:00 hrs (November 24): Geocoding and introduction to replication projects</u>

- Contents
    - Geocoding non-spatial data in R
    - Introduction to replication projects and automatic reports with R
- Key readings
    - AidData Research and Evaluation Unit. (2017). Geocoding Methodology, Version 2.0.1. Williamsburg, VA: AidData at William & Mary.
    - BenYishay, A., Rotberg, R., Wells, J., Lv, Z., Goodman, S., Kovacevic, L., Runfola, D. 2017. Geocoding Afrobarometer Rounds 1-6: Methodology & Data Quality. Williamsburg, VA: AidData at William & Mary.
- Lab activities
    - Geocoding cities and other points on a map
- Assignment 6 issued

Week 7

<u>Friday 11:00 – 12:30 hrs (December 1): Replication – Night lights</u>
- Contents
  - Measuring welfare in developing countries
  - Advantages and pitfalls of night lights
- Key readings
  - (replication) Alesina, A., S. Michalopoulos, and E. Papaioannou. 2016. Ethnic Inequality. *Journal of Political Economy*, 124(2), 428–488.
  - (background) Henderson, J. V., A. Storeygard, and D. N. Weil. 2012. Measuring Economic Growth from Outer Space. *American Economic Review*, 102(2), 994–1028.
- Lab activities
  - Calculating ethnic and spatial inequality in Afghanistan
  - Creating Voronoi (Thiessen) polygons and "virtual" homelands

<u>Friday 13:30 – 15:00 hrs (December 1): Replication – Night lights (continued)</u>

Week 8

<u>Friday 11:00 – 12:30 hrs (December 8): Replication – Slave trade</u>
- Contents
  - Slave trade as an obstacle to the formation of larger ethnic groups and trust
  - Distances as an instrument and controls
- Key readings
  - (replication) Nunn, N. 2008. The Long-Term Effects of Africa's Slave Trades. *Quarterly Journal of Economics*, 123(1), 139–176.
  - (background) Nunn, N., and L. Wantchekon. 2011. The Slave Trade and the Origins of Mistrust in Africa. *American Economic Review*, 101(7), 3221–3252.
- Lab activities
  - Calculating the overland distance to slave trade centers
  - Calculating sailing distances to slave trade centers

<u>Friday 13:30 – 15:00 hrs (December 8): Replication – Slave trade (continued)</u>

Week 9

<u>Friday 11:00 – 12:30 hrs (December 16): Replication – Conflict or Road Networks (your choice, if time allows)</u>
- Contents
  - The persistence, location, and causes of conflict in Africa
  - Reliability of geolocated conflict data **OR**
  - Transport modes and the spatial distribution of economic activity
  - Economic shocks and spatial diffusion
- Key readings
  - (replication) Berman, N., Couttenier, M., Rohner, D., and M. Thoenig. 2017. This Mine Is Mine! How Minerals Fuel Conflicts in Africa. *American Economic Review*, 107(6), 1564–1610. **OR**

- (replication) Storeygard, A. 2016. Farther on down the Road: Transport Costs, Trade and Urban Growth in Sub-Saharan Africa. *Review of Economic Studies,* 83(3), 1263–1295.
- Lab activities
  - Assigning conflict events and other point data to grid cells, and
  - Panel fixed effect regressions with spatial HAC standard errors in R **OR**
  - Calculating measures of network connectivity and shortest paths

Friday 13:30 – 15:00 hrs (December 16): Replication – Conflict or Road Networks (continued)

Week 10

Friday 11:00 – 12:30 hrs (December 22): Leftovers for this year (online)

Friday 13:30 – 15:00 hrs (December 22): Independent work on replications

Week 11

Friday 11:00 – 12:30 hrs (January 12): Lab hours with TA

Week 12

Friday 11:00 – 12:30 hrs (January 19): Lab hours with TA

Week 13

Friday 11:00 – 12:30 hrs (January 26): Lab hours with TA

Week 14

Friday 11:00 – 15:00 hrs (February 2): Presentations of replications

Week 15

no class