MODULE 9: Spatial Statistics in Epidemiology and Public Health
Lecture 1: Introduction (You Are Here)

Lance Waller and Howard Chang

Department of Biostatistics and Bioinformatics
Rollins School of Public Health
Emory University

Welcome!

- Spatial Statistics in Epidemiology and Public Health
- Spatial statistics? In Epidemiology?
- *Where* something happens may offer insight into *what* is happening and *why*.
- Some ideas from statistics and epidemiology transfer over directly, some don't.
- We want *spatial* answers to spatial questions, and we want *statistical* answers to statistical questions.
- What do *spatial statistical* questions (and answers look like)?

Course logistics

- Your instructors (both from Emory University):
  - Lance A. Waller
  - Howard Chang
- Framework: Modules and (short) labs in 90 minute blocks.
- Course materials on Github

Monday schedule

- 8:30 - 10: Lectures 1 and 2: Introduction, Maps, Mapping, GIS (Lance)

  10-10:30: Break
- 10:30-noon: Lecture 2: Maps, Mapping, Spatial Questions and Answers (Lance)

  Noon-1:30: Lunch (on your own)
- 1:30-3: Lecture 3: Area data (EDA, clustering) (Howard)

  3-3:30: Break
- 3:30-5: Lecture 4: Disease mapping (CAR, SAR) (Howard)
- 5-7pm: Networking night and poster session (R. Randall Rollins Room P01 – Plaza Level)

Tuesday schedule

- 8:30 - 10: Lecture 5: Spatial regression and slippery slopes (Lance)

  10-10:30: Break

- 10:30-noon: Lecture 6: Gaussian processes (Howard)

  Noon-1:30: Lunch (on your own)

- 1:30-3: Lecture 7: Point processes (Lance)

  3-3:30: Break

- 3:30-5: Lecture 8: Space-time models (Howard)

Wednesday schedule

- 8:30 - 10: Lecture 9: Disease ecology (Lance)

  10-10:30: Break
- 10:30-noon: Lecture 10: Multivariate processes (Howard)

References: Spatial Statistics

- Waller, L.A. and Gotway, C.A. (2004) *Applied Spatial Statistics for Public Health Data*. Wiley.
- Lawson, A.B. (2023) *Using R for Bayesian Spatial and Spatio-Temporal Health Modeling*. CRC.
- Banerjee, S., Carlin, B.P, and Gelfand, A.E. (2015) *Hierarchical Modeling and Analysis for Spatial Data*. CRC.
- Blangiardo, M., Cameletti, M. (2015) *Spatial and Spatio-temporal Bayesian Models with R-INLA*. Wiley.
- Diggle, P.J., Giorgi, E. (2019) *Model-based Geostatistics for Global Public Health: Methods and Applications*. CRC.

References: Mapping

- Monmonier, M. (2018) *How to Lie with Maps, Third Edition*. Chicagor.
- MacEachren, A. (1995) *How Maps Work*. Guilford.
- Koch, T. (2017) *Disease Maps*. Chicago.
- Koch, T. (2005) *Cartographies of Disease*. ESRI Press.
- Walker, K. (2023) *Analyzing U.S. Census Data: Methods, Maps, and Models in R*. CRC. Also available online at walker-data.com/census-r/index.html
- Moraga P (2020) *Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny*. Boca Raton: Chapman & Hall/CRC.
- Andrienko N et al. (2020) *Visual Analytics for Data Scientists*. Springer.
- Waller LA (2024) Maps: A statistical view. *Annual Review of Statistics and Its Applications*

Scoping

- What spatial data do you work with?
- What spatial questions are of interest?

Lectures and Labs and Course Materials

- Each session will have a lecture component and a (short) lab component.
- Natalie Olson, our TA (natalie.olson@emory.edu), created a Github repository for the course with all lecture notes, lab data, and lab Rmd (R markdown) files.
- There are pdf files of the lecture notes, there are pdf markdown files for the labs, and Rmd (R markdown) files for the labs.
- Go to https://github.com/lance-waller-lab/2024-SISMID-Spatial-Statistics

Getting the files

- Go to the green "Code" button and download as a zip file.
- Save the zip file to a directory of your choice, then unzip the file. You will see a folder of lectures, a folder for data, etc.
- IMPORTANT: You will also see a file 2024-SISMID.Rproj this will pull in all of the lab code and data as a project.
- If you are unfamiliar with R/RStudio, you can focus on reading through the lab pdfs to see examples of what you can do.
- If you are familiar with R/RStudio, you can try out the Rmd code.

Opening as an R Project in RStudio

- Open RStudio.
- Under the File menu, select "Open Project".
- Navigate to the directory where you stored the downloaded folder.
- Open the 2024-SISMID.Rproject file.
- Under the 'Files' tab in the lower right quadrant of RStudio, you'll see folders for each of the lectures.
- Inside each of these folders will be the pdf and Rmd files for that lecture/lab combination.

Some R notes

- R is a very popular, open-source environment, and there are many packages for different parts of spatial analysis.
- Some of the lab R code is based on old (in some cases, very old) packages but we have tested them for your use.
- There are many new packages for working with different types of data (e.g., geographic information system shapefiles).
- Walker (2023) gives a great entry for mapping and working with U.S. Census data: walker-data.com/census-r/index.html
- Moraga (2020) provides illustrations of many mapping features and several mapping packages in R.
- Andrienko et al. (2020) gives some very nice data visualization advice for mapping (especially Chapters 9-10).

A changing spatial geography: sp and sf

- https://r-spatial.org/ provides the latest on setting standards for spatial data handling in R.
- sp is the former standard for spatial processing and many R functions relied on it.
- sf is the new standard (and is different from sp, but improved in many ways).
- There was a goal of independence from sp by October 2023, but there are some legacy packages supporting it and some old packages that require it.
- It is important to know this is going on when working in spatial statistics and using R.
- Some of the details are in the weeds, but they do influence the trees!

Let's get the repositories set up

- Follow the directions above.
- Download the zip file.
- Unload the zip file.
- Open RStudio.
- Open the Project.
- Make sure you can open the Rmd files.

Helpful References: Spatial Statistics

- Waller, L.A. and Gotway, C.A. (2004) *Applied Spatial Statistics for Public Health Data*. Wiley.
- Lawson, A.B. (2023) *Using R for Bayesian Spatial and Spatio-Temporal Health Modeling*. CRC.
- Banerjee, S., Carlin, B.P, and Gelfand, A.E. (2015) *Hierarchical Modeling and Analysis for Spatial Data*. CRC.
- Blangiardo, M., Cameletti, M. (2015) *Spatial and Spatio-temporal Bayesian Models with R-INLA*. Wiley.
- Diggle, P.J., Giorgi, E. (2019) *Model-based Geostatistics for Global Public Health: Methods and Applications*. CRC.

Helpful References: Mapping

- Monmonier, M. (2018) *How to Lie with Maps, Third Edition*. Chicagor.
- MacEachren, A. (1995) *How Maps Work*. Guilford.
- Koch, T. (2017) *Disease Maps*. Chicago.
- Koch, T. (2005) *Cartographies of Disease*. ESRI Press.
- Walker, K. (2023) *Analyzing U.S. Census Data: Methods, Maps, and Models in R*. CRC. Also available online at walker-data.com/census-r/index.html
- Moraga P (2020) *Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny*. Boca Raton: Chapman & Hall/CRC.
- Andrienko N et al. (2020) *Visual Analytics for Data Scientists*. Springer.
- Waller LA (2017) Mapping in Public Health. In *Mapping Across Academia*, Brunn, S.D. and Dodge, M., eds. Dordrecht: Springer.