

Department/Interdisciplinary Programme

Academic Year 2015 - 2016

Advanced Quantitative Methods for Spatial and Big Data Analysis

SP067 – Autumn 2015-16 - 6 ECTS

Wednesday 12:15-14:00 (S3)

> PROFESSOR

Karsten Donnay
karsten.donnay@graduateinstitute.ch
+41 22 908 59 42

Office Hours:
MdP-P2-632
Wednesday 14:00-16:00,
by appointment only

> ASSISTANT

Alessandro Regio
alessandro.regio@graduateinstitute.ch
+41 22 908 45 29

Office Hours:
MdP-P2-601
TBD,
by appointment only

Course Description

This course provides a basic introduction to two areas of quantitative research methodology with increasing relevance for social scientists. In the first module, students are familiarized with spatially disaggregate data, spatial polygons, and point data. As a concrete application, the course will provide students with a hands-on introduction to evidence-driven computational modeling, a technique that makes it possible to run simulations on representative geographic topologies of real world cases. A second module of the course will focus on the analysis of Big Data. In particular, how data gleaned from social media sites may be used for substantive social science research. Emphasis will be placed on the limitations of current research and general difficulties in using data from various social media sites. The course will serve as a point of departure for students (and faculty) interested in applying these methodologies to their own research. A good understanding of basic quantitative methods is a prerequisite; knowledge of advanced quantitative methods is an asset.

Syllabus

Course Requirements

Requirement 1: Attendance in class is required and students are expected to engage with the recommended readings and/or online resources in preparation for each session. It is essential that you come to class prepared and actively participate.

Requirement 2: Students will be required to complete a written exam at the end of the term. The exam will focus on conceptual rather than technical skills. No course material, computers or any other kind of aids will be needed or allowed. The exam will take place in the last session of the term and take 90 minutes.

Course Evaluation

Performance in the course depends both on active participation and performance in the final exam. Evaluation will be based on:

- | | |
|---------------------------------|-----|
| 1. Attendance and Participation | 20% |
| 2. Final Exam | 80% |

Course Material

The following textbooks are recommended as general (technical) reference works for spatial data analysis and Big Data. Recommended readings and/or online resources for each session are provided with stable links in the course schedule below.

- Longley, Paul A., Michael F. Goodchild, David J. Maguire, and David W. Rhind. (2015) *Geographic Information Systems and Science* 4th Ed. [Wiley](#).
- Bivand, Roger S., Edzer Pebesma and Virgilio Gómez-Rubio. (2013). *Applied Spatial Data Analysis with R* [Springer](#).
- McGovern, Tim (ed.). (2015). [Big Data Now: 2014 Edition](#). O'Reilly Media.

The books on GIS and spatial data analysis are on reserve at the library. The book on Big Data is a compilation of blog posts that covers a wide range of topics. Its scope is beyond that of this class but it is a good resource to browse and read up on specific topics. It can be downloaded for free via the link provided (the previous 2013 edition is available [here](#)).

Overview of Class Sessions

The course is divided into two modules, one focusing on spatial data analysis, the other on Big Data. Sessions are structured around specific topics. The focus of the course is on concepts but technical aspects are introduced and discussed where necessary. Technical demonstrations are done in *R* and their code will be made available.

Part 1: Spatial Data Analysis

1. Introduction
2. Basic Concepts
3. Spatial Data Structures
- 4.* Practical Session - Handling Spatial Data in R
5. Data Coding and Quality
6. Methodological Challenges
- 7.* Standard Methodologies for Spatial Data Analysis
- 8.* Advanced Methodologies for Spatial Data Analysis
9. Example Application - Evidence Driven Modeling

Part 2: Big Data

10. Introduction - What is Big Data?
- 11.* Handling and Processing Big Data
12. Methodological Challenges and Problems
13. Example Application – Using Twitter to Analyze Political Discourse

* lectures that include demonstrations/examples in R

Course Schedule with Recommended Readings and Online Resources¹

Module 1: Spatial Data Analysis

Session 1: Introduction

Wednesday, September 16

- Gleditsch, Kristian S., and Nils B. Weidmann. (2012). [Richardson in the Information Age: Geographic Information Systems and Spatial Data in International Studies](#). *Annual Review of Political Science* 15: 461-481.
- Raleigh, Clionadh, Frank Witmer, and John O'Loughlin. (2010). [A Review and Assessment of Spatial Analysis and Conflict: The Geography of War](#). in R. Denmark (ed) *The International Studies Encyclopedia Vol X*. Wiley-Blackwell. pp. 6534-6553.
- Tam Cho, Wendy K., and James G. Gimpel. (2012). [Geographic Information Systems and the Spatial Dimensions of American Politics](#). *Annual Review of Political Science* 15: 443-460.

Session 2: Basic Concepts

Wednesday, September 23

- ArcGIS Resource Center. [What is GIS?](#)
- GIS Wiki. [Datum](#).
- ArcGIS Resource Center. [Map Projections](#) and [Geographical Coordinate Systems](#).

Session 3: Spatial Data Structures

Wednesday, September 30

- Geospatial Innovation Facility. [GIS Data Types](#).
- GeoTIFF. [FAQ](#) and [Format Specification](#).
- Shapefile. [Description](#) and [Formal Definition](#).
- ArcGIS Resource Center. [What is Raster Data](#) and [Raster Coordinate Systems](#).

Session 4: Practical Session – Handling Spatial Data in R

Wednesday, October 7

- Rodriguez-Sanchez, Francisco. (2013). [Spatial data in R: Using R as a GIS](#).
- Bivand, Roger. (2015). [CRAN Task View: Analysis of Spatial Data](#).
- Lovelance, Robin and James Cheshire. (2015). [Introduction to Visualizing Spatial Data in R](#).
- Fowler, Chris. (2011). [GIS Workshops – Spatial R](#).

Session 5: Data Coding and Quality

Wednesday, October 14

- Eck, Kristine. (2012). [In Data We Trust? A Comparison of UCDP GED and ACLED Conflict Events Datasets](#). *Cooperation and Conflict* 47(1): 124-141.
- Salehyan, Idean. (2015). [Best Practices in the Collection of Conflict Data](#). *Journal of Peace Research* 52(1): 105-109.
- Weidmann, Nils B. (2013). [The Higher the Better? The Limits of Analytical Resolution in Conflict Event Datasets](#). *Cooperation and Conflict* 48(4): 567-576.
- Weidmann, Nils B. (2015). [On the Accuracy of Media-based Conflict Event Data](#). *Journal of Conflict Resolution* 59(6): 1129-1149.

Session 6: Methodological Challenges

Wednesday, October 21

- Openshaw S. 1984. [The Modifiable Areal Unit Problem](#). Geobooks, Norwich, England.
- Witmer, Frank D., Dark, Shawna J., and Danielle Bram. (2007). [The Modifiable Areal Unit Problem \(MAUP\) in Physical Geography](#). *Progress in Physical Geography* 31(5): 471-479.
- Gotway, Carol A., and Linda J. Young. (2002). [Combining Incompatible Spatial Data](#). *Journal of the American Statistical Association* 97(458): 632-648.
- Witmer, Frank D. (2015). [Remote Sensing of Violent Conflict: Eyes From Above](#). *International Journal of Remote Sensing* 36(9): 2326-2352.

¹ The instructor reserves the right to modify the course syllabus during the semester depending on student progress.

Session 7: Standard Methodologies for Spatial Data Analysis

Wednesday, October 28

- Anselin, Luc. (2001). [Spatial Econometrics](#). In *A Companion to Theoretical Econometrics* (ch. 14) edited by Badi H. Baltagi (Blackwell Publishing), p. 310-330.
- LeSage, James, and R. Kelley Pace. (2009). [Introduction to Spatial Econometrics](#). Taylor & Francis.
- Elhorst, J. Paul. (2014). [Spatial Econometrics – From Cross-Sectional Data to Spatial Panels](#). Springer.

Session 8: Advanced Methodologies for Spatial Data Analysis

Wednesday, November 4

- Linke, Andrew M., and John O'Loughlin. (2015). [Reconceptualizing, Measuring, and Evaluating Distance and Context in the Study of Conflicts: Using Survey Data from the North Caucasus of Russia](#). *International Studies Review* 17(1): 107-125.
- Linke, Andrew M., Sebastian Schutte, and Halvard Buhaug. (2015). [Population Attitudes and the Spread of Political Violence in Sub-Saharan Africa](#). *International Studies Review* 17(1): 26-45.
- Sebastian Schutte, and Karsten Donnay. (2014). [Matched wake analysis: Finding causal relationships in spatiotemporal event data](#). *Political Geography* 41: 1-10. ([R package](#))

Session 9: Example Application: Evidence Driven Modeling

Wednesday, November 11

- Ravi Bhavnani, Karsten Donnay, Dan Miodownik, Maayan Mor, and Dirk Helbing. (2014). [Group Segregation and Urban Violence](#). *American Journal of Political Science* 58(1): 226-245
- Weidmann, Nils B., and Idean Salehyan. (2013). [Violence and Ethnic Segregation: A Computational Model Applied to Baghdad](#). *International Studies Quarterly* 57(1): 52-64.

Module 2: Big Data

Session 10: Introduction – What is Big Data?

Wednesday, November 18

- Dutcher, Jenna. (2014). [What is Big Data?](#) *UC Berkeley Data Science Blog*.
- Press, Gil. (2014). [12 Big Data Definitions: What's Yours?](#) *Forbes Blog*.
- Manovich, Lev. (2012). [Trending: The Promises and the Challenges of Big Social Data](#). *Debates in the Digital Humanities*, edited by Matthew K. Gold. The University of Minnesota Press.
- Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall Van Alstyne. (2009). [Computational Social Science](#). *Science* 323(5915): 721-723.

Session 11: Handling and Processing Big Data

Wednesday, November 25

- Atz, Ulrich. (2013). [11 Tips on How to Handle Big Data in R](#). *Open Data Institute Blog*.
- Lockwood, Glenn. (2014). [Conceptual Overview of Map-Reduce and Hadoop](#). *Blog Post*.
- Jacobs, Bill. (2015). [Using Hadoop with R: It Depends](#). *Blog Post*.
- Penchikala, Srinii. (2015). [Big Data Processing in Apache Spark – Part 1: Introduction](#). *InfoQ Article*.
- Venkataraman, Shivaram. (2015). [Announcing SparkR: R on Spark](#). *Databricks Blog Post*. ([R package](#))

Session 12: Methodological Challenges and Problems

Wednesday, December 2

- Bollier, David (2010). [The Promise and Peril of Big Data](#). *The Aspen Institute Communications and Society Program*.
- Cate, Fred H. (2014). [The Big Data Debate](#). *Science* 346(6211): 818-818.
- Lazer, David, Ryan Kennedy, Gary King, and Alessandro Vespignani. (2014). [The Parable of Google Flu: Traps in Big Data Analysis](#). *Science* 343(6176): 1203-1205.
- Lazer, David. (2015). [The Rise of the Social Algorithm](#). *Science* 348(6239): 1090-1091.
- Ulfelder, Jay. (2015). [The Myth of Comprehensive Data](#). *Blog Post*.
- Weller, Nicholas, and Kenneth McCubbins. (2014). [Raining on the Parade: Some Cautions Regarding the Global Database of Events, Language and Tone Dataset](#). *Blog Post*.

Session 13: Example Application – Using Twitter to Analyze Political Discourse

Wednesday, December 9

Viktoria Spaiser, Thomas Chadeaux, Karsten Donnay, Fabian Russmann, and Dirk Helbing. (2014). [Social Media and Regime Change: The Strategic Use of Twitter in the 2011-12 Russian Protests](#). SSRN doi:10.2139/ssrn.2528102.

Session 14: Final Exam

Wednesday, December 16