

3-14 JULY 2017

# BARCELONA summer school - OF DEMOGRAPHY -



## *Introduction to R & Data Visualization*

Organized by the Centre d'Estudis Demogràfics (CED)  
at the Universitat Autònoma de Barcelona (UAB)

### INSTRUCTORS

**Juan Galeano** [jgaleano@ced.uab.es](mailto:jgaleano@ced.uab.es)

Center for Demographic Studies (CED), Autonomous University of Barcelona

**Tim Riffe** [riffe@demogr.mpg.de](mailto:riffe@demogr.mpg.de)

Max Planck Institute for Demographic Research, Rostock, Germany

**Francisco Villavicencio** [villavicencio@imada.sdu.dk](mailto:villavicencio@imada.sdu.dk)

Max-Planck Odense Center on the Biodemography of Aging, University of Southern Denmark, Odense, Denmark

### COURSE OUTLINE

The aim of this course is to provide an intensive introduction to R for researchers wanting to dive deep into this powerful and open source statistical software. The course is divided in two different modules, one for each week, covering 3 major strengths of R: Statistical analysis, data visualization, and spatial analysis.

The first module Introduction to R and Basic Statistics will introduce R, data manipulation, and basic statistical analysis. The second module Data Visualization and Spatial Analysis will first focus on tools for data visualization, introducing the ggplot2 R package. Next, the students will learn how to handle spatial data in R, elaborate digital cartography, and perform basic spatial analyses.

No previous knowledge in R is required to attend the course. Students are expected to bring their own laptops with R and RStudio installed. Both modules will be taught in English.

The course is limited to 18 participants selected on a competitive basis. Priority will be given to researchers early-career (Master and PhD students) but applicants from more advanced stages are also welcome.



## SCHEDULE AND ORGANIZATION

The course will be held at the Center for Demographics Studies (CED), located on the Campus of the Autonomous University of Barcelona, Bellaterra, Spain, from 3 to 14 July 2017. Lectures will be divided into two daily sessions. Morning sessions (from 10 to 13:30, with a 30 minutes break) will be devoted to theoretical lectures, combined with practical exercises. Afternoon sessions (from 15:00 to 17:00, except Fridays) will be computer labs in which students, under the supervision and help of the instructors, will be asked to work on different R exercises. In addition to the teaching sessions, several social activities are planned, including a one-day guided tour of the city of Barcelona on Saturday, 8 July and a farewell barbecue on Friday, 14 July.

		MORNING 10-13:30h.	AFTERNOON 15-17h.
<b>WEEK 1</b> 3-7 JULY	MONDAY	R & Basic Stats	Computer Lab
	TUESDAY	R & Basic Stats	Computer Lab
	WEDNESDAY	R & Basic Stats	Computer Lab
	THURSDAY	R & Basic Stats	Computer Lab
	FRIDAY	R & Basic Stats	
	SATURDAY	Social activities	
	SUNDAY		
<b>WEEK 2</b> 10-14 JULY	MONDAY	DATAVIZ	Computer Lab
	TUESDAY	DATAVIZ	Computer Lab
	WEDNESDAY	Spatial R	Computer Lab
	THURSDAY	Spatial R	Computer Lab
	FRIDAY	Spatial R	

### MODULE 1 INTRODUCTION TO R AND BASIC STATISTICS

#### Session 1 (Monday)

- Welcome reception by Albert Esteve, Director of the Center for Demographic Studies.
- Introduction to R and RStudio.
- Using the editor: main characteristics of RStudio, packages, help options.
- Data handling: import/export data to/from R.
- Basic R objects: vectors.
- Basic mathematical functions.

#### Session 2 (Tuesday)

- Data types in R.
- Basic R objects: matrices, arrays, data frames, lists.
- Introduction to data simulation in R: the `runif()`, `rnorm()` and `rt()` family functions.

#### Session 3 (Wednesday)

- Basic statistical functions: mean, variance, standard deviation, covariance, IQR, etc.
- Tables.
- Introduction to R plotting: histograms, boxplots, scatter plots, and lines.
- Exporting graphics.

#### Session 4 (Thursday)

- Hypothesis testing (baisc theory), p-values.
- Hypothesis testing in R (ANOVA): analysis and interpretation.
- Linear models in R: Measures of goodness of fit (p-values, coefficient of determination, total sum of squares), and interpretation of the output of the built-in R functions.

#### Session 5 (Friday)

- Flow control: conditional execution (if) and for-loops.
- The apply family functions.
- Creating your own functions in R.

## MODULE 2 DATA VISUALIZATION AND SPATIAL ANALYSIS

### Session 6 (Monday): Data Visualization

1. Primitive plot commands.
2. Fine-tuning aesthetics.
3. Pyramids.
4. Surfaces.

### Session 7 (Tuesday): Data Visualization

1. Introducing ggplot2 and its underlying idea: the grammar of graphics.
2. Basic plots in ggplot2.
3. Data structure: long vs wide format.
4. Facets.
5. Scales and its attributes (name, limits, breaks, labels).
6. Continuous versus discrete color scales.
7. Themes.

### Session 8 (Wednesday): Spatial R

1. Read shapefiles into R.
2. General manipulation of spatial objects.
3. Univariate Class Intervals.
4. Color palettes.
5. Thematic maps (I).

### Session 9 (Thursday): Spatial R

1. Conversion between projection systems.
2. The ggmap package.
3. Thematic maps (II).

### Session 10 (Friday): Spatial R

1. Spatial Statistics.
2. Measures of spatial segregation and population diversity: The OasisR package.
3. Neighborhood Matrix.
4. Spatial autocorrelation: Global and Local Indicators of Spatial Autocorrelation (LISA).

## APPLICATION

Candidates should apply by an email to [bssd@ced.uab.es](mailto:bssd@ced.uab.es). Please write “Application to BSSD” in the subject of your email. You are also asked to attach the following documents (in English): 1) Your Curriculum Vitae (free format) and 2) A one-page cover letter, summarizing your research interests and how they are related to the course. Please state also if you have any previous experience with R or any other statistical software (e.g. STATA, SPSS).

The deadline apply is **Wednesday, 15 March 2017**. Students will be informed of their acceptance by 31 March 2017.

## COURSE FEE AND ACCOMMODATION

The course has a fee of 75 Euro, to be paid upon acceptance. This fee will include participation in social activities that will be organized during the course.

Students are expected to pay their own transportation and living costs. The CED can arrange affordable housing on Campus (around 20 Euro per night): <https://vilauniversitaria.uab.cat/en/>. If interested, please state so in your application email.

Please do not hesitate to ask for further information by sending an email to [bssd@ced.uab.es](mailto:bssd@ced.uab.es).

---

## ABOUT THE INSTRUCTORS



**Juan Galeano** [jgaleano@ced.uab.es](mailto:jgaleano@ced.uab.es)

Center for Demographic Studies (CED), Autonomous University of Barcelona

Juan Galeano holds a PhD in Demography from the Center for Demographic Studies (CED) and the Autonomous University of Barcelona (UAB). Master in Demography from the European Doctoral School of Demography (EDSD), Master in Territorial and Population Studies from CED and UAB, and BA in Sociology from the University of Barcelona (UB). His current research focuses on the spatial and demographic consequences of the settlement of migrants in Spain since the beginning of the XXI century, mainly increasing population diversity and residential segregation.



**Tim Riffe** [riffe@demogr.mpg.de](mailto:riffe@demogr.mpg.de)

Max Planck Institute for Demographic Research, Rostock, Germany

Tim Riffe is a research scientist at the Max Planck Institute for Demographic Research. His theoretical work focuses on population renewal and temporal relationships over the life course. His empirical work uses original methodological approaches to study relationships between longevity and health in ageing populations, based on both administrative and survey data.



**Francisco Villavicencio** [villavicencio@imada.sdu.dk](mailto:villavicencio@imada.sdu.dk)

Max-Planck Odense Center on the Biodemography of Aging, University of Southern Denmark, Odense, Denmark

Francisco Villavicencio has a background in mathematics and geography, and in 2012-2013 he attended the European Doctoral School of Demography. He is currently working at the University of Southern Denmark, and his research interests include the development of methods to deal with sparse demographic data (Bayesian inference, agent-based modeling). In the last two years he has taught in several courses of R and statistics at the Department of Mathematics and Computer Science of that university.