ANT 291: Data analysis using R

Damien Caillaud

Fall 2017

E-mail: dcaillaud@ucdavis.edu
Office Hours: Thursday 10am-12pm
Office: 216 Young Hall

Web: **canvas.ucdavis.edu**
Lectures: M 1:10-3:00pm, 210 Young Hall
Labs: W 2:20-3:00pm, SCILAB 2020

Course Description

ANT 291 is a PhD-level introduction to data management, data visualization and statistical inference. Lectures will present theoretical explanations and practical solutions to some of the major statistical problems graduate students in biology, anthropology and related fields are likely to encounter. More specifically:

- Options to collect/store/manage research data (Excel, relational databases, cloud-based storage...)

- Using the R language to manipulate data, make pretty graphs and run statistical models

- Basic linear modelling: understanding and visualizing fixed effects, interactions, significance testing

- Generalized Linear Models (GLMs): binomial, negative binomial, Poisson, zero-inflated Poisson models, etc...

- Generalized Linear Mixed-Effects Models (GLMMs): understanding "frequentist" and Bayesian definitions of random effects...

- Choosing among models

- Refining linear models to take into account temporal autocorrelation and spatial autocorrelation

- Basic Geographical Information System (GIS) analyses using R

- Basic network analysis using R

- Basic multivariate analysis

- Bayesian vs. frequentist statistics

- Multicollinearity, AIC, multiple-testing, autocorrelation, model checking, P-hacking and other words that will give you headaches...

Lectures will all be based on concrete examples. The goal of the course is not to make you become an expert in statistical theory. Lectures will not include a lot of equations. Be ready for quite a lot of R coding, though. You will run scripts I will write for you and will create your own scripts. Luckily, the R language is one of the simplest computer languages. You don't need to have spent your summer reading R books or following online tutorials to take this course.

Getting ready for the course

If you own a laptop, don't hesitate to bring it to class. R is pretty easy to install on Windows, MacOS and Linux machines: https://cran.r-project.org/.
If you have a Mac, you will probably be pretty happy with the default R script editor. Unfortunately the Windows version of R does not come with a nice color-coded editor so you will probably want to use RStudio.
   If you want to read online R tutorials to complement the lectures, here are a few useful links:

- https://cran.r-project.org/doc/manuals/r-release/R-intro.html

- https://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf

- http://www.burns-stat.com/documents/books/the-r-inferno/

RBlogger is also a fantastic resource both for beginners and experts.

Readings

Dozens of books about R and data analysis are available. Most of them are pretty good. The only book I'll advise you to read is this Bayesian statistics book written by Richard McElreath. I would however advise you to read it right *after* you've taken ANT 291. It will help you re-memorize all the concepts I will have talked about.

Home assignments and exams

Each week, I will give post a homework problem on Canvas. I will post the solutions the week after. You are welcome to work together on these assignments as long as you make sure you understand every bit of code you're running. Some of the problems will be easy, but others will be more challenging. So keep in mind that you are not expected to be able to answer every question! I will give more details about what is the most important during class.
   The final exam will be a take-home exam. It will include simple questions and will be intended to verify that you are able to use R to solve statistical problems. You will have to work

alone on this final exam. You will be allowed to use the Internet and books, but you won't be allowed to ask another person.

Grading Policy

The Course is letter-graded.

- Attendance will account for 30% of your final grade.

- Homework assignments (done/not done) will account for 30% of your final grade.

- The final exam will account for 40% of your final grade

Depending on the distribution of the final grades, I may curve the grades upwards.

Academic Dishonesty Policy

You are expected to abide by the Code of Academic Conduct. This set of policies forbids cheating, plagiarism, and other forms of academic dishonesty.

Disabilities Policy

Please contact me by email if you need any special accommodation.

Phone policy

Please turn your phone off when you come to the class room.

Schedule

Week 00, 09/25 - 09/29:   Introduction (Wednesday only)

Week 01, 10/02 - 10/06:   Basic R coding: variable types, tables, functions, loops, git...

Week 02, 10/09 - 10/13:   Linear modelling in R using lm() and glm()

Week 03, 10/16 - 10/20:   Mixed-effects models using lme4

Week 04, 10/23 - 10/27:   Bayesian statistics

Week 05, 10/30 - 11/03:   Model selection (AIC, DIC, WAIC, BIC...)

Week 06, 11/06 - 11/10:   Time series

Week 07, 11/13 - 11/17:   Spatial data analysis

Week 08, 11/20 - 11/24:   Multivariate analyses

Week 09, 11/27 - 12/01:   Network data analysis

Week 10, 12/04 - 12/08:   Review session