IR 413: Applied Data Science for International Relations 2

Spring 2018
Professor: Benjamin Graham
Office: VKC 357B
E-mail: benjamin.a.graham@usc.edu
Web Site: http://dornsife.usc.edu/graham
Office Hours: By Appointment: Wednesdays 1:30-4:30PM
https://benjamingraham.youcanbook.me/

Stats Consulting Office Hours: Times TBD.
Sign up for statistical computing help here: https://statsconsulting.youcanbook.me/

****
This syllabus will be updated throughout the semester. New versions of the syllabus will always be announced in lecture, and students are responsible for ensuring they have the latest version. The most up to date version will always be on my website.
****

This course provides an introduction to data analysis and the graphical presentation of statistical relationships using R (http://cran.r-project.org). This is the second part of a data science sequence following IR 412. The goals of this course are to enable students to manage and clean data, including reshaping and merging datasets; to understand the intuition behind regression analysis; and create publication-grade figures in R that graphically display regression results.

This course centers around two large projects.

In the first, each student contributes a new dataset to the International Political Economy Data resource, which is a compilation dataset maintained by the Security and Political Economy Lab at USC. Each student identifies a useful country year data source that is not currently included in the IPE Data Resource, and adds that dataset to the resource. This requires soliciting permission from the dataset authors, downloading the data, cleaning and merge-prepping the data, merging it into the IPE data resource, and writing a new entry for the codebook.

In the second project, each student creates a publication-grade figure that graphically presents the regression results from an actual working paper or book chapter written by an SIR faculty member. Professor Graham will match faculty and students, and faculty who solicit figures for their work are encouraged to attend the final sessions of the course when students present their work.
Pre-requisites:
IR 210, IR 211, and either IR 412, IR 307: Introduction to Data Analysis, or equivalent. Permission to enroll will also be granted at the discretion of the professor to students who have completed IR210 and IR 211 and have sufficient prior experience working in R.

Assignments & Grading
World Development Indicators Practice Assignment: 15%

Data Management Project: 30%
Each student must contribute a new country-year dataset to the IPE data resource. See above.

Graphical Presentation of Regression Results Project: 40%
Students will produce publication-grade figures that graphically present the results of the statistical analysis in an actual working paper or book chapter authored by an SIR faculty member. Many of these figures will become part of actual published research. See above.

Required Doing in Swirl: 10%
Students are required to hand in a log file of the required doing at the beginning of class on the day for which the required doing is assigned.

Final Presentation: 5%

Online Resources
Swirl (Learn R, in R): http://swirlstats.com/
Cookbook for R: http://www.cookbook-r.com/
Quick R: http://statmethods.net/

Required Text:
Kosuke Imai, Quantitative Social Science: An Introduction

Recommended Texts:
Adler, J. (2010), R in a Nutshell, O’Reilly.

Statement for Students with Disabilities
Any student requesting academic accommodations based on a disability is required to register with Disability Services and Programs (DSP) each semester. A letter of verification for approved accommodations can be obtained from DSP. Please be sure the letter is delivered to me as early in the semester as possible. DSP is located in STU 301 and is open 8:30 a.m.–5:00 p.m., Tuesday through Friday. The phone number for DSP is (213) 740-0776.
Statement on Academic Integrity
USC seeks to maintain an optimal learning environment. General principles of academic
honesty include the concept of respect for the intellectual property of others, the
expectation that individual work will be submitted unless otherwise allowed by an
instructor, and the obligations both to protect one’s own academic work from misuse by
others as well as to avoid using another’s work as one’s own. All students are expected to
understand and abide by these principles. Scampus, the Student Guidebook, contains the
Student Conduct Code in Section 11.00, while the recommended sanctions are located in
Appendix A: http://www.usc.edu/dept/publications/SCAMPUS/gov/. Students will be
referred to the Office of Student Judicial Affairs and Community Standards for further
review, should there be any suspicion of academic dishonesty. The Review process can
be found at: http://www.usc.edu/student-affairs/SJACS/.

Lecture Topics, Assignment Dates, and Readings

Week 1: January 12
Course Introduction and Planning

Week 2: January 19
Lecture 1: Data Cleaning I
Prior to class:
- Required Viewing: 046; 047; 069; 078 (http://www.twotorials.com/)
- Required Reading: “Subsetting Data”
  (http://statmethods.net/management/subset.html)
- Required Reading: “Sorting Data”
  (http://statmethods.net/management/sorting.html)
- Recommended Reading: “Data Processing with dplyr & tidyr”
  (https://rpubs.com/bradleyboehmke/data_wrangling)
- Assignment 1 will be handed out (WDI Practice Exercise).

Week 3: January 26
Lab 1: Data Cleaning I
Prior to class:
- Required viewing: 042; 028; 054; 079 (http://www.twotorials.com/)
- Required Doing: Swirl: Complete course “Getting and Cleaning Data”:
  1(Manipulating Data with dplyr); 2(Grouping and Chaining with dplyr)

In class exercise:
  1) dplyr package:
     - select(); filter(); arrange(); mutate(); summarise(); group_by()

Week 4: February 2
Lecture 2: Data Cleaning II
****** Assignment 1 Due (WDI Practice Exercise)******
Prior to class:
• Required Reading: “Introducing tidyr”  
  (http://blog.rstudio.org/2014/07/22/introducing-tidyr/)  
• Optional Reading: “Restructuring Data”  (http://www.cookbook-r.com/Manipulating_data/)

Week 5: February 9  
Lab 2: Data Cleaning II
Prior to class:
• Required Reading: “Converting data between wide and long format”  
  (http://www.cookbook-r.com/Manipulating_data/Converting_data_between_wide_and_long_format/)  
• Required Reading: “Reshaping Data”  
  (http://statmethods.net/management/reshape.html)  
• Required Doing: Swirl: Complete course “Getting and Cleaning Data”: 3(Tyding Data with tidyr)

In class exercise:
1) reshape2 package:
   • melt( ); cast( )  
2) tidyr package:
   • gather( )  
   • separate( )  
   • spread( )

Week 6: February 16  
Lecture 3: Simple Linear Regression (intuition)  
• Required Reading:

Week 7: February 23  
Lab 3: Simple Linear Regression (intuition)  
****Project 1 is due at the beginning of class****  
• In Class:
  o Swirl: Complete “Regression Models” course: Lesson 2(Residuals); 3(Least Squares Estimation)

Week 8: March 2  
Lecture 4: Multiple Regression (intuition)  
• Required Reading:

Week 9: March 9  
Lab 4: Multiple Regression (intuition)  
In Class:  
Swirl: Complete the following lessons: 5 - Introduction to Multivariable Regression; 6 - MultiVar Examples; 7 - MultiVar Examples2; 7 - MultiVar Examples3
Week 10: March 23
Lecture 5: Graphical presentation of regression results
   Scatter plots with lines of best fit
   Ladder plots
   Varying the leads plots
   Regression discontinuities & splines

****Project 2 assignments are made****

Week 11: March 30
Lab 5: Graphical presentation of regression results
   In Class: Ladder plot exercise

***Hand drawn sketch of final projects is due***

Week 12: April 6
Lab 6: Collaborative Work Session

Week 13: April 13
Lab 7: Collaborative Work Session

Week 14: April 20
Final Project Presentations & Critique

Week 15: April 27
Final Project Presentations & Critique

Final Exam Week: Revised final projects are due on the day of the final.