Syllabus for Spring 2019 Data Analysis and Statistics for Political Science Research

Time: MW 2:40pm-3:55pm Location: 309 Havemeyer Hall

Michael D. Parrott <u>mp3675@columbia.edu</u> Office Hours: IAB 509f, Friday, 3-5

Teaching Assistants: Jorge Mangonnet jgm2159@columbia.edu Office Hours: IAB 736, Thursday, 9:15-11:15

Seungyup Shin ss5370@columbia.edu Office Hours: IAB 734, Tuesday, 2-4

Office Hours: IAB 731, Monday, 10-12

Jose Gomez-Espinal

jq3341@columbia.edu

Course Goals

This course examines the basic methods of data analysis and statistics that political scientists use in quantitative research that attempts to make causal inferences about how the political world works. The same methods apply to other kinds of problems about cause and effect relationships more generally. The course will provide students with extensive experience in analyzing data and in writing (and thus reading) research papers about testable theories and hypotheses.

How to conduct new research and how to interpret research are key aspects of (political) science. Research allows us to pose novel questions, construct fresh theories, and provide new evidence about how the world works. However, before we start doing research, we have to learn how it works.

This course teaches you how to think scientifically about politics. It is designed to help students develop their analytical abilities and statistical reasoning. It will provide an introduction to research design and the use of quantitative methods in political science. In other words, you will learn how to formulate questions and theories about politics and then test them in a rigorous manner.

Students can expect to have exposure to actual and hypothetical cases, as well as hands-on practice with research tools (e.g., statistical software program R). Thus, this course will combine teaching the principals of statistics as used in political science with data analysis—that is, it will involve both theory and practice.

Prerequisites and Software

We do not expect students to have a mathematical background beyond high school algebra. We also assume that students have little or no experience using computers for data analysis. *We will be using R in this class*. R is free and works with any computer. We will be coding with R using the R-Studio IDE. R-Studio is a free interface

for R that assists coders in a number of helpful ways.

I will walk through the set up for R and R-Studio early in the course. The Digital Social Science Center (DSSC) in Lehman Library has this software installed on all computers. If you have any difficulties setting up R or R-Studio on your computer please use the computers in the DSSC. You will find similar assistance and software in the Digital Science Center (DSC) in the Science & Engineering Library.

You will often learn by doing. Students will benefit from bringing a laptop to each class, though this is not strictly required. To execute code for in-class examples you will need access to a computer with the latest version of R and R-Studio.

Course Expectations

<u>Attendance and Class Participation</u>. Class attendance is required and is necessary for your success in the course. We will begin taking attendance using sign-in sheets in the second week of class. You are allowed to miss up to four (4) days of class without penalty. Missing more than four (4) days of class will require you to speak with the instructor, and it will affect your course grade.

Exams. We will have two exams that will ask you to apply what you have learned in lectures and homework assignments.

<u>Homework Assignments</u>. Students will have six homework assignments due throughout the semester. They will be based on writing up the results of concepts and code learned during the lectures. Specific instructions, format and deadlines will be given as the semester progresses. All assignments should be turned in through the assignments tab on the course website.

<u>Plagiarism and Academic Dishonesty</u>: Students must do all their work within the boundaries of acceptable academic norms. See the Academic Honesty page of the CU website regarding college policy on plagiarism and other forms of academic dishonesty -

<u>http://www.columbia.edu/cu/history/ugrad/main/handbook/academic_honesty.html</u>. Students found guilty of plagiarism or academic dishonesty will be subject to appropriate disciplinary action, which may include reduction of grade, a failure in the course, suspension or expulsion.

<u>Disability Services:</u> In order to receive disability-related academic accommodations for this course, students must first be registered with their school Disability Services (DS) office. Detailed information is available online for both the *Columbia and Barnard* registration processes. For this course, students are not required to have testing forms or accommodation letters signed by faculty. The Instructor section of the form has already been completed and does not need to be signed by the professor. The student must complete the Student section of the form and submit the form to Disability Services. Master forms are available in the Disability Services office or online: https://health.columbia.edu/services/testing-accommodations.

Late Assignments. Students will lose points for handing in late assignments, at the discretion of the instructor and teaching assistants.

<u>Use Piazza for Q & A</u>. This term we will be using Piazza for class discussion. The system is highly catered to getting you help fast and efficiently from classmates, the TA, and myself. Rather than emailing questions to the teaching staff, I encourage you to post your questions on Piazza. If you have any problems or feedback for the

developers, email team@piazza.com.

Find our class page at the below link or via the link provided on the left hand side of the course website: https://piazza.com/columbia/spring2019/polsun3704_001_2019_1dataanalysisstatspolres/home

Textbooks. The following books will help you further your understanding of the material:

- Paul M. Kellstedt and Guy D. Whitten: The Fundamentals of Political Science Research, Second Edition, 2013. (other editions are also okay to purchase)
- Hadley Wickham and Garret Grolemund: R for Data Science, First Edition, 2017. This will help with the usual tasks of data import, data management, and some visualization. (available **for free** online <u>here</u>)
- Matthew Salganik: Bit by Bit: Social Research in a Digital Age. 2017. (available for free online here)

The course will closely follow Kellstedt and Whitten (K&W in the below course outline) for content about research design, statistics, and data analysis. R for Data Science (RDS) will be our go to book for learning how to work with data using R. Bit by Bit will allow us to expand our conversation towards the end of the class into cutting edge data collection and analysis as we consider social science methods in a digital age.

<u>Additional Materials</u>. Other articles and materials will be distributed via Courseworks that cover additional topics in more depth.

<u>Grade Distribution</u>. The distribution of your final grade will be calculated as follows:

Two Exams = 30% Homework Assignments = 60% Attendance and Participation = 10%

<u>Changes</u>: There may be adjustments in the scheduling of assignments, exams, and classrooms. Changes will be posted through Courseworks along with other announcements.

Course Outline, Readings, and Assignments

Class 1 (January, 23rd). Introduction; The Scientific Study of Politics.

Reading Assignments
K&W Ch. 1, "Facing the Truth" Economist Article

Class 2 (January, 28th). Software Infrastructure: *Getting Started with R and R-Studio* Note: HW 1 Assigned (Code: Loading data into R and interacting with it)

Reading Assignments

Install R and R studio, Work through Chptrs. 1,2,3 and 5 of Data Camp's Intro to R tutorials (Chptrs 4 and 6 are useful, but optional)

Class 3 (January, 30th). The Art of Theory Construction.

Reading Assignments

Class 4 (February 4th). Intro to Data Manipulation in R.

Note: **HW 1 is due**. HW 2 Assigned (Code: Interacting with data in R using the Tidyverse)

Reading Assignments	
RDS Ch. 5, 9 (short), and 10	

Class 5 (February, 6th). Causality and Causal Inference, Elements of Research Design

Reading Assignments	
K&W Ch. 3-4	

Class 6 (February, 11th). Definition and Measurement of Concepts, Levels of Measurement, and Framing Hypotheses Note:

Reading Assignments	
K&W Ch. 5, Hypothesis Handout	

Class 7 (February, 13th). Understanding Univariate Variable distributions, Making Simple Bivariate Analyses (With examples in R). **HW 2 is due.**

Reading Assignments	
K&W Ch. 6	

Class 8 (February, 18th). Foundations of Statistical Inference.

Note: HW 3 Assigned (Paper #1: Explore variables in a dataset, make and evaluate hypotheses)

Reading Assignments	
K&W Ch. 7	

Class 9 (February, 20th). Catch Up and Review Concepts and Code

Reading Assignments	
RDS Ch. 7	

Class 10 (February, 25th). Bivariate Hypothesis Testing - Comparing two Sample Proportions and Chi Square Tests (With examples in R). Note: **HW 3 is due.**

Reading Assignments	
K&W Ch. 8	

Class 11 (February, 27th). Bivariate Hypothesis Testing, Comparing Two Means (With examples in R). Note: HW 4 Assigned (Paper #2: Analyze bivariate relationships and evaluate statistical significance)

Reading Assignments	
K&W Ch. 8	

Class 12 (March 4th). Bivariate Hypothesis Testing - Correlation (With examples in R)

Reading Assignments	
K&W Ch. 8	

Class 13 (March, 11th). Review Concepts and Code for Test Note: **HW 4 is due.**

Reading Assignments	
No readings this week	

Class 14 (March, 13th). Mid-Term Exam

Spring Break

Class 15 (March, 25th). Matrix Operations for Ordinary Least Squares Linear Regression

Reading Assignments	
No readings this week	

Class 16 (March, 27th). Simple Linear Regression

Reading Assignments	
K&W Ch. 9	

Class 17 (April, 1st). Simple Linear Regression (cont'd)

Reading Assignments	
No further readings this week	

Class 18 (April, 3rd). Multivariate Linear Regression

Note: HW 5 Assigned (Paper #3: Testing hypotheses using Bivariate and Multivariate Regression)

Reading Assignments	
K&W Ch. 10	

Class 19 (April, 8th). Multivariate Linear Regression (cont'd)

Reading Assignments	
No further readings this week	

Class 20 (April, 10th). Categorical IVs in OLS, Interaction terms Note: **HW 5 is due.**

Reading Assignments	
K&W Ch. 11	

Class 21 (April, 15th). Multivariate Linear Regression Examples in Research

Reading Assignments	
K&W Ch. 12	

Class 22 (April, 15th). Multivariate Linear Regression Examples in Research

Reading Assignments	
K&W Ch. 12	

Class 23 (April, 17th). Common Problems with OLS Models, Violations of Gauss Markov Assumptions

Reading Assignments	
K&W Ch. 12	

Class 24 (April, 22nd). Common Problems with OLS Models, Violations of Gauss Markov Assumptions Note: HW 6 Assigned (Paper #4: Putting it all together. Write a 10 to 12 page paper that lays out hypotheses, runs bivariate analysis, runs multivariate model, and then observes residuals to speculate about potential violations of OLS assumptions on a dataset provided via the course website.) **Paper due on May 16th.**

Reading Assignments	
K&W Ch. 12	

Class 25 (April, 24th). Special Topic: Social Science Research in a Digital Age

Reading Assignments	
Bit by Bit, Chs. 1 & 2	

Class 26 (April, 29th). Special Topic: Social Science Research in a Digital Age

Reading Assignments	
No further readings	

Class 27 (May, 1st). Review Concepts and Code for Test

Reading Assignments	
No further readings	

Class 28 (May, 6th). Final Exam