# DPI 610: Data Science for Politics

Harvard Kennedy School | Spring 2021 | Monday 12:00pm-1:15pm | Zoom

## **Benjamin Schneer**

Assistant Professor of Public Policy Harvard Kennedy School Office: Taubman 364 Office Hours: W, 12:00pm-2:00pm or by appointment Email: benjamin\_schneer@hks.harvard.edu

Teaching Fellow: Soichiro Yamauchi Office Hours: Th, 1:00pm-3:00pm or by appointment Email: syamauchi@g.harvard.edu

Course Assistant: Rasim Alam Office Hours: Th, 9:00am-11:00am or by appointment Email: mrasim@hks.harvard.edu

## **Course Description**

Decision-making in politics is now more informed by data than ever before. Data analysis guides voter targeting by campaigns, predictions about election outcomes, and critical policy decisions made by government officials — to name just a few important areas touched by the revolution in the availability and use of data. This course covers key areas of politics transformed in recent years by data science, and it introduces fundamental tools of data science through applications to politics. The course takes a problem-driven approach, covering background and academic literature on a topic, learning a relevant data science tool or method, and then applying it to real-world data. A primary goal of the course is to give students an opportunity to develop data analysis skills relevant for working in politics, including writing and implementing code in statistical software packages; through applications students will gain experience with data wrangling/cleaning/formatting, regression, prediction, visualization, unstructured data, and spatial data.

The course is organized by topic into four modules:

- Campaigns
- Redistricting
- Measuring and Analyzing Opinions
- Forecasting Elections

In each module, students will have the opportunity to develop experience and skills working with related data through group programming exercises and homework assignments.

## **Course Learning Goals**

- 1. The course will help you build skills in data science, primarily with regard to programming in **R**.
- 2. The course will help you read, interpret and critique empirical social science. Some of the material in the course draws upon academic journal articles. Our coverage of this material will help you think about translating rigorous research into applicable lessons for practitioners.

## **Course Policies**

- *Enrollment*: The course is open to all HKS students and students in other programs at Harvard. Visitors may take course with permission of instructor.
- *Prerequisites*: There are no official prerequisites for this course; however, I recommend some experience with basic statistics, such as from taking API-201.
- *Participation & Attendance*: Students should show up to class having done the assigned reading and/or programming exercise, ready to listen, ask relevant questions and discuss. In the case of a missed class due to an emergency, you are responsible for getting notes and other relevant information from a classmate.
- *E-mail and Communication Policy*: I encourage questions and comments through e-mail and at office hours. To receive a response, please e-mail me using your Harvard e-mail address.
- Office Hours: Office hours are Wednesday from 1:00pm-3:00pm.
- *Grading Issues*: If you feel that there was an error in the grading of an assignment, you may submit a re-grading request to me. This should be done only in extraordinary circumstances. If I am not able to clarify your concerns in person, I ask that you submit a written request for a re-grade along with a memo outlining where an error in the grading occurred. Note that a re-grade may result in a higher or a lower grade when the assignment is reconsidered.
- *Americans With Disabilities Act*: Students with disabilities who need accommodation may contact the HKS Student Disability Coordinator for assistance.
- Academic Honesty: You must adhere to the Kennedy School's policies regarding appropriate academic practices. The academic code is available <u>here</u> for your reference and on the Canvas site. Academic honesty in writing is of particular relevance for this course. You must cite

sources and provide proper attribution for ideas and words that are not your own, as outlined in the HKS academic code.

## **Course Requirements**

The final course grade consists of the following items:

- **Class Participation and Class Programming Exercises)** (20%): I ask that you participate actively in class if you are able to attend. The class benefits greatly from discussion and good discussions are generally facilitated by everyone having prepared their thoughts in advance. In addition, a second part of course participation will consist of weekly or bi-weekly group programming exercises. If you are in a location that makes live attendance impossible, then participation in the programming exercises and discussion on Canvas covers this portion of the course grade.
- Homework Assignments (20%): Assignments graded by the Course Assistant on a zero/check-minus/check/check-plus scale. We will drop one missed or low-scoring assignment.
- **Group Project** (30%): Students will perform some analysis in order to make a clear recommendation on a pressing matter of public policy. Specific assignments will be distributed later in the course. Half of the assignment consists of writing a memo and the other half requires the students to give a presentation.
  - Presentations: April 26 or April 28.
  - Memo Due: May 3.
- Take Home Final Exam (30%): A final "take home" assignment will be administered based on the material covered in the course. More details to follow.

PhD students interested in taking this course for credit may do so by completing a longer research paper in lieu of the final exam and group project.

Final grades for HKS students will be assigned according to the Dean's Recommended Grade distribution (available <u>here</u>).

## **R** Resources

The first step to learning R is to install the software. Specifically, install R (https://cran.rproject.org/) and RStudio (https://rstudio.com/products/rstudio/download); we will also use Rstudio.cloud, an online platform that is useful for distributing assignments and group programming exercises. Directions to set up an account will be provided later.

Probably the best way to learn R is through learning by doing. But as a first introduction to R and as a reference, parts (I recommend chapters 1, 2, 4, 19, 20, 21, and 27) of the following book are a good resource:

Wickham, Hadley, and Garrett Grolemund. 2016. *R for data science: import, tidy, transform, visualize, and model data*. O'Reilly Media. Available free online at https://r4ds.had.co.nz/.

## **Reading Materials**

Readings will be posted on the course website. Additionally, this syllabus is a *living document*. Changes to the reading list will occur as we progress through the semester as may homework assignment and due dates.

## **Meeting Time and Location**

This class meets on Mondays from 12:00pm–1:15pm online. Additional instruction will occur asynchronously through short video recordings and group programming exercises. The suggested time for completing Group Programming Exercises will be Wednesday afternoons; course staff will be available from 12pm–6pm to assist on Wednesdays. On the schedule, group programming exercises are thus specified on Wednesdays.

## **Course Schedule**

Jan. 22 (Fri.) Shopping Day: Introduction

#### Campaigns

Jan. 25 (Mon.) Exploring Voter Data and Making Predictions

Nickerson, David W., and Todd Rogers. 2014. "Political campaigns and big data." *Journal of Economic Perspectives* 28 (2): 51–74.

Hersh, Eitan D. 2015. *Hacking the electorate: How campaigns perceive voters*. Cambridge University Press, chapters 1, 4 (selection), and 5.

#### Jan. 27 (Wed.) Class Exercise: Intro to R and RStudio; Voter File Intro

*Assignment to Complete Beforehand:* Install R and RStudio; Complete online Tutorials (Linked on Course Website)

#### Feb. 1 (Mon.) Voter Targeting

Malchow, Hal. 2008. "Political targeting." *Washington, DC: Campaigns and Elections*, chapters 3 and 4.

Bond, Robert M, Christopher J Fariss, Jason J Jones, Adam DI Kramer, Cameron Marlow, Jaime E Settle, and James H Fowler. 2012. "A 61-million-person experiment in social influence and political mobilization." *Nature* 489 (7415): 295–298.

Kafka, Peter. 2019. "Facebook's political ad problem, explained by an expert." https://www.vox.com/ recode/2019/12/10/20996869/facebook-political-ads-targeting-alex-stamos-interview-opensourced.

Smets, Kaat, and Carolien van Ham. 2013. "The embarrassment of riches? A meta-analysis of individual-level research on voter turnout." *Electoral Studies* 32 (June): 344–359.

Feb. 3 (Wed.) Class Exercise: Prediction Models for Turnout and Persuasion

HW 1 Assigned (Due 2/19/2021)

Feb. 8 (Mon.) Evaluating Impact: Potential Outcomes and Turnout

Green, Donald P., and Alan S. Gerber. 2019. *Get out the vote: How to increase voter turnout*. Brookings Institution Press, chapters 1–3, 5 & 7.

Feb. 10 (Wed.) Class Exercise: Randomized Experiments

Feb. 15 (Mon.) NO CLASS: President's Day

Feb. 17 (Wed.) No Class Exercise

HW 1 Due 2/19/2021

Feb. 22 (Mon.) Evaluating Impact: Persuasion

Broockman, David, and Joshua Kalla. 2016. "Durably reducing transphobia: A field experiment on door-to-door canvassing." *Science* 352 (April): 220–224.

Kalla, Joshua L., and David E. Broockman. 2018. "The Minimal Persuasive Effects of Campaign Contact in General Elections: Evidence from 49 Field Experiments." *American Political Science Review* 112 (February): 148–166.

Kalla, Joshua L, and David E Broockman. 2020. "Reducing Exclusionary Attitudes through Interpersonal Conversation: Evidence from Three Field Experiments." *American Political Science Review*: 16.

HW 2 Assigned (Due 3/8/2021)

Feb. 24 (Wed.) Class Exercise: Immigration Experiment

Mar. 1 (Mon.) NO CLASS: HKS Wellness Day

Mar. 3 (Wed.) No Class Exercise

#### Redistricting

Mar. 8 (Mon.) Redistricting and the Ecological Inference Problem

Stewart, Charles Haines. 2001. Analyzing congress. Norton New York, pp. 205-226.

Into the Thicket: A Redistricting Starter Kit for Legislative Staff. 2019 http://www.ncsl.org/research/ redistricting/into-the-thicket-a-redistricting-starter-kit-for-legislative-staff.aspx.

Robinson, W. S. 1950. "Ecological Correlations and the Behavior of Individuals." *American Sociological Review* 15 (3): 351–357.

Gelman, Andrew, David K. Park, Stephen Ansolabehere, Phillip N. Price, and Lorraine C. Minnite. 2001. "Models, assumptions and model checking in ecological regressions." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 164 (January): 101–118.

#### **Optional Further Reading:**

Greiner, D. James, and Kevin M. Quinn. 2009. "RxC ecological inference: bounds, correlations, flexibility and transparency of assumptions." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 172 (1): 67–81.

#### HW 2 Due

Mar. 10 (Wed.) Class Exercise: Spatial Data and Visualization

**Mar. 15 (Mon.)** Measuring Partisan Gerrymandering: Traditional Techniques and Simulations

Royden, Laura, and Michael Li. 2017. Extreme Maps. Technical report Brennan Center for Justice.

King, Gary, and Robert X Browning. 1987. "Democratic representation and partisan bias in congressional elections." *The American Political Science Review*: 1251–1273.

Stephanopoulos, Nicholas O., and Eric M. McGhee. 2015. "Partisan Gerrymandering and the Efficiency Gap." *University of Chicago Law Review* 82 (2): 831–900.

#### **Optional Further Reading:**

Duchin, Moon. 2018. "Gerrymandering Metrics: How to Measure? What's the Baseline?" (January): 6.

Chen, Jowei, and Jonathan Rodden. 2015. "Cutting Through the Thicket: Redistricting Simulations and the Detection of Partisan Gerrymanders." *Election Law Journal: Rules, Politics, and Policy* 14 (December): 331–345.

Chen, Jowei. 2017. "The Impact of Political Geography on Wisconsin Redistricting: An Analysis of Wisconsin's Act 43 Assembly Districting Plan." *Election Law Journal: Rules, Politics, and Policy* 16 (December): 443–452.

#### HW 3 Assigned (Due 3/29/2021)

Mar. 17 (Wed.) Class Exercise: Measuring Gerrymanders

#### Measuring and Analyzing Opinions

Mar. 22 (Mon.) Sampling, Collecting and Analyzing Survey Data

Groves, Robert M., Floyd J. Fowler Jr, Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. 2009. *Survey Methodology*. 2 edition ed. Hoboken, N.J: Wiley, chapters 3 and 4 (Optional Reading: chapter 2).

Groves, Robert M., Floyd J. Fowler Jr, Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. 2009. *Survey Methodology*. 2 edition ed. Hoboken, N.J: Wiley, chapter 10.

Gelman, Andrew, and Thomas C. Little. 1998. "Improving on Probability Weighting for Household Size." *Public Opinion Quarterly* 62 (3): 398.

Final Project Task 1 Assigned: Form group and Send Paragraph Describing Project to Instructor (by 4/5/2021)

Mar. 24 (Wed.) Class Exercise: Surveys: Weighting, Raking, and Poststratification

Mar. 29 (Mon.) Polling in the 2016 and 2020 Elections

Cohn, Nate. 2017. "A 2016 Review: Why Key State Polls Were Wrong About Trump." *The New York Times* (May).

Kennedy, Courtney, Mark Blumenthal, Scott Clement, Joshua D. Clinton, Claire Durand, Charles Franklin, Kyley McGeeney, Lee Miringoff, Kristen Olson, Douglas Rivers, Lydia Saad, G. Evans Witt, and Christopher Wlezien. 2018. "An Evaluation of the 2016 Election Polls in the United States." *Public Opinion Quarterly* 82 (March): 1–33.

Panagopoulos, Costas. 2021. "Polls and Elections Accuracy and Bias in the 2020 U.S. General Election Polls." *Presidential Studies Quarterly* 51 (1): 214–227.

Enns, Peter K. 2021. "Revisiting the 'gold standard' of polling: new methods outperformed traditional ones in 2020." https://medium.com/3streams/revisiting-the-gold-standard-of-polling-newmethods-outperformed-traditional-ones-in-2020-451650a9ba5b.

### HW 3 Due

### Project Description Paragraph Due

Mar. 31 (Wed.) NO CLASS: HKS Wellness Day

## Forecasting Elections

## Apr. 5 (Mon.) Poll Aggregation Models and Fundamental Models

Campbell, James E., Helmut Norpoth, Alan I. Abramowitz, Michael S. Lewis-Beck, Charles Tien, James E. Campbell, Robert S. Erikson, Christopher Wlezien, Brad Lockerbie, Thomas M. Holbrook, Bruno Jerôme, Véronique Jerôme-Speziari, Andreas Graefe, J. Scott Armstrong, Randall J. Jones, and Alfred G. Cuzán. 2017. "A Recap of the 2016 Election Forecasts." *PS: Political Science & Politics* 50 (April): 331–338.

Jackman, Simon. 2005. "Pooling the polls over an election campaign." *Australian Journal of Political Science* 40 (December): 499–517.

Fair, Ray. 2011. *Predicting presidential elections and other things*. Stanford University Press, chapters 1–4.

## HW 4 Assigned (Due 4/19/2021)

## Apr. 7 (Wed.) Class Exercise: Predicting Election Outcomes I

Apr. 12 (Mon.) Synthesized Models and Prediction Markets I

Lauderdale, Benjamin E., and Drew Linzer. 2015. "Under-performing, over-performing, or just performing? The limitations of fundamentals-based presidential election forecasting." *International*  Journal of Forecasting 31 (July): 965–979.

Linzer, Drew A. 2013. "Dynamic Bayesian Forecasting of Presidential Elections in the States." *Journal of the American Statistical Association* 108 (March): 124–134.

Rothschild, David. 2009. "Forecasting Elections: Comparing Prediction Markets, Polls, and Their Biases." *Public Opinion Quarterly* 73 (January): 895–916.

#### **Optional Further Reading:**

Fisman, Raymond, and Eric Zitzewitz. 2019. "An Event Long-Short Index: Theory and Applications." *American Economic Review: Insights* 1 (September): 357–372.

## Apr. 14 (Wed.) Class Exercise: Predicting Election Outcomes II

Apr. 19 (Mon.) Prediction Markets II, Non-Representative Data and MRP

Rothschild, David. 2015. "Combining forecasts for elections: Accurate, relevant, and timely." *International Journal of Forecasting* 31 (July): 952–964.

Wang, Wei, David Rothschild, Sharad Goel, and Andrew Gelman. 2015. "Forecasting elections with non-representative polls." *International Journal of Forecasting* 31 (July): 980–991.

#### **Optional Further Reading:**

Ghitza, Yair, and Andrew Gelman. N.d. "Voter Registration Databases and MRP." *Political Analysis*. Forthcoming.

#### HW 4 Due

Apr. 21 (Wed.) No Class Exercise: Work on Group Projects

#### Conclusions

Apr. 26 (Mon.) Student Presentations

Apr. 28 (Wed.) Student Presentations