Quantitative Political Analysis I

From Data to Modeling

New College of Florida • Fall 2026

Instructor
Young Joon Oh

Email
yoh@ncf.edu

Office
Social Science 205

Office Hours
Tuesday and Thursday, 10:00–11:00 a.m.;
Tuesday, 2:00–3:00 p.m.; and by appointment

Meeting Time
Monday and Wednesday, 11:00 a.m.–12:20 p.m.

Location
Academic Center 237

Canvas
Announcements, readings, public data, annotated R scripts, and project materials

Course Overview

Why do some communities vote at higher rates than others? Are younger citizens less trusting of government? Do wealthier countries tend to be more democratic? Questions like these appear in campaigns, policy debates, news stories, and senior project — and anyone can offer an answer. But an answer built on data can persuade, because it can withstand scrutiny: how is the dataset put together? What does each variable actually measure? Which comparisons are meaningful? How uncertain are the results? What conclusions does the evidence actually support?

This course teaches that process from beginning to end: understanding, managing, and modeling data to connect a political question to evidence. Working with real public data, students first learn to read a dataset and its documentation, then to describe and visualize what the data contain, and finally to carry that understanding into regression models that estimate how a political outcome moves with an explanatory factor. By the end of the semester, each student completes a manageable data analysis: locating and documenting a public dataset, preparing selected variables in R, exploring the data with statistics and visualizations, estimating and checking regression models, and communicating the findings in a short paper and presentation. You will leave with a compact empirical research paper, a reproducible R script, and a presentation of your findings — a foundation you can adapt for a senior project, or later methods courses.

QPA I is a foundation, not a survey. It deliberately covers a set of tools — description, visualization, basic inference, and regression modeling — and teaches each one carefully, on real data. That preparation lets you read quantitative political claims critically and carry out a focused, defensible analysis of your own; it is also the natural first step toward the optional later methods courses.

Political question → Public data → Data management → Description and visualization → Regression modeling → Argument

Guiding principle: Start with the question, not the tool. Be a social scientist, not a regressionist: the question, the data, and the intended claim should determine the method.

Where This Course Leads

The political-methods sequence is organized around research problems that grow progressively trickier — problems whose answers require weighing more considerations at once. The later courses are optional and build directly on this one.

Course	Central questions	Main capability
QPA I: From Data to Modeling	How can we use data to answer a research question? What patterns can we discover in a dataset? How do we move from concepts to variables, and from variables to a statistical model? How should we analyze and interpret the results?	Turn a research question into a clear and reproducible data analysis.
QPA II: Beyond the Linear Model	Which model best fits the research question and the data? What should we do when the outcome is categorical? Is it reasonable to analyze California and Florida voters as if their different social and political contexts did not matter? What if people's decisions are shaped by socioeconomic groups to which they belong?	Choose, compare, and interpret generalized linear and multilevel models.
Advanced Statistical Modeling : Modeling Political Complexity	How should we analyze data when the same people, states, or countries are observed repeatedly over time? How do we model trends, temporal dependence, and change? How can we produce and evaluate forecasts? Can prior belief and new evidence be combined in analysis ?	Analyze longitudinal and time-series data, produce and evaluate forecasts, and apply selected Bayesian ideas.

No prior statistics or programming is required. Students should be comfortable with percentages, negative numbers, simple equations, and reading tables and graphs.

How the Course Works: Statistics, Data, and R

QPA I has no separately scheduled laboratory section. Regular meetings combine explanation, brief calculations, interpretation, discussion, and guided work with public data. The pattern throughout the course is *learn, then do*: a concept is developed on the board first, and the same class then produces it in R. Occasionally, a class meeting may be held in a computer classroom instead of the usual room; any such change will be announced on Canvas.

- The instructor provides annotated R scripts and demonstrates each step of the workflow in class as it becomes relevant. All scripts and data used in class are posted on Canvas, so every demonstration can be reproduced exactly at home.
- **Students install R and RStudio on their own computers**, following H2. The essential practice is simple: attend class, then reproduce what we did on your own computer the same week. This practice is the real preparation for the proposal, the presentation, and the final paper.
- R syntax is not tested through memorization; students are, however, responsible for reproducing

and explaining their own project workflow.

Required Materials

- **Main textbook:** Diez, David M., Mine Çetinkaya-Rundel, and Christopher D. Barr. *OpenIntro Statistics*, 4th ed. Available free from OpenIntro.
- **Recommended for R and regression:** Hanck, Arnold, Gerber, and Schmelzer. *Introduction to Econometrics with R* — free online at econometrics-with-r.org.
- **Recommended:** Agresti, Alan. *Statistical Methods for the Social Sciences*, recent ed., for additional social-science examples and depth.
- **Instructor handouts and annotated R scripts,** posted on Canvas.
- **Calculator:** a non-programmable scientific calculator (such as the TI-30XIIS or Casio fx-300ES Plus) for quizzes and the midterm.
- **R and RStudio:** free.

Learning Outcomes

By the end of the course, students should be able to:

1. Translate a political question into observable cases, a defensible unit of analysis, and measurable variables.
2. Locate a credible public dataset, read its documentation, and cite the data correctly.
3. Prepare and document selected variables in R, including transparent decisions about missing values.
4. Describe and visualize political data with appropriate statistics, tables, and graphics.
5. Explain sampling uncertainty and interpret confidence intervals, tests, and group comparisons.
6. Estimate and cautiously interpret regression models and read residual patterns as warning signs.
7. Communicate a reproducible empirical argument in a concise paper, tables, figures, and a short presentation.

Evaluation

Component	Weight
Participation and preparation	10%
Four quizzes (best three count)	20%
Midterm examination	25%
Project proposal	5%
Final presentation	5%
Data analysis paper	35%

There is no final examination. The proposal, presentation, and paper together constitute the project.

Participation and Preparation

Participation is assessed through attendance, preparation, and constructive engagement with class activities. The best preparation for the final project is simple: attend class, then reproduce what we did on your own computer the same week.

Quizzes

There are four noncumulative quizzes of about 20 minutes each; only the best three scores count. Two fall before the midterm and two after. Quizzes are closed-note (**no cheat sheet**) and combine statistical reasoning, method choice, interpretation, and short calculations; memorized R syntax is never tested. A **calculator** is required.

Quiz	Date (11-11:20am)
Quiz 1	Wednesday, September 16
Quiz 2	Wednesday, October 14
Quiz 3	Wednesday, November 18
Quiz 4	Wednesday, November 25

A missed quiz ordinarily becomes the dropped quiz. Routine make-up quizzes are not offered.

Midterm Examination

The midterm examination is held on **Monday, November 2** and covers Units 1–9 (all material through October 28). The 80-minute exam emphasizes interpretation, method selection, short calculations, and statistical reasoning. A **calculator** is required, and **one letter-size handwritten cheat sheet** is permitted. R syntax and software operation are not tested.

Project Proposal

Due **Friday, November 6, at 11:59 p.m. on Canvas** (5%). A short document (2–3 pages) containing: the research question; a description of the data — the source and producing organization, what the full dataset covers, and which portion of it (cases and variables) the analysis will use — with the unit of analysis identified; and a preliminary look at those variables using the tools of the first half of the course: a descriptive-statistics table and at least one visualization, briefly interpreted. The proposal shows that the data are in hand, understood, and feasible before the regression models are estimated on them ; students may refine the question's wording, variable construction, or model after receiving feedback, provided that major changes are discussed with the instructor.

Final Presentation

During the final instructional week (5%), each student gives a concise presentation: the regression models estimated on the proposal's data, the key results with a table and a visualization, and their interpretation. Details and refinements belong in the paper — and because the paper is due after the presentations, students may use questions and feedback from the presentation to improve the final submission. Presentation guidelines, order, and time limits are included in the Project Guide (PG) on Canvas; the per-student time limit depends on final enrollment.

Data Analysis Paper

The final paper is a **6–8 page paper, including tables and figures** (35%), that answers a focused political question with a public dataset approved through the proposal. The paper is not simply a regression exercise: students first understand and describe their data, then estimate two regression models, check them, and explain how the pieces support a defensible argument.

Students submit the paper and the R script used to produce the evidence. The project is evaluated on the quality of the question, data decisions, statistical reasoning, visual communication, interpretation, transparency, and writing. The Project Guide (PG) is posted on Canvas by October 19. The **final paper is due Friday, December 11, at 11:59 p.m. on Canvas**.

Essential Course Policies

- **Missed work:** A missed quiz ordinarily becomes the dropped quiz. Midterm or presentation rescheduling, or a paper extension, requires a documented illness, family emergency, approved travel, or another circumstance recognized by College policy. Contact the instructor promptly.
- **Academic integrity and AI:** Follow the New College Academic Honor Code. AI tools may not be used on quizzes or the midterm. For the project, AI tools may be used to explain error messages or clarify instructor-provided R code, with brief acknowledgment. Students may not use AI to generate the project's research question, make data or modeling decisions, produce substantive interpretations, or draft submitted prose or slides. Students remain responsible for verifying and explaining every element of their work.
- **Support and communication:** Students seeking accommodations should contact the Accessible Learning Center and the instructor early. Course materials and updates appear on Canvas; students should check Canvas and NCF email regularly.

Course Schedule

Four topical instructor handouts (H1–H4) carry the material the textbooks do not cover. The schedule follows the official Fall 2026 calendar.

PART I — Statistical Reasoning and Public-Data Literacy

Unit 1 — Political Questions and the Data Workflow

Mon, Aug 17

What can political data tell us?

Main: Course syllabus

Political questions; evidence and claims; the semester workflow; the data-analysis project; the political methods sequence.

Wed, Aug 19

Cases, variables, populations, and samples

Main: OpenIntro §§1.2 and 1.3.1

Observations and variables; units of analysis; populations and samples; description, association, prediction, and causation.

Unit 2 — Measurement and Data Documentation

Mon, Aug 24

From political concepts to variables

Main: OpenIntro §1.2 ; Agresti §2.1 ; Handout H1

Operationalization; levels of measurement; validity, reliability, and measurement error.

Wed, Aug 26

Reading a dataset before analyzing it

Main: OpenIntro §1.2; Handout H1

Variable names, labels, values, units, documentation, missing-value codes, and analysis samples.

Unit 3 — Describing and Visualizing Data

Mon, Aug 31

Describing distributions and groups

Main: OpenIntro §§2.1 and 2.2.1

Frequencies and proportions; mean and median; standard deviation and interquartile range; comparing distributions.

Wed, Sep 2

Visual evidence and misleading displays

Main: OpenIntro §§2.2.1, 2.2.3, and 2.2.6; Handout H1

Bar charts, histograms, boxplots, grouped comparisons, and misleading displays (examples on slides); how the day's graphics look when R draws them, following H1's appendix — the mechanics of R come later in the course.

Mon, Sep 7 *Labor Day — no class.*

Unit 4 — Probability and Uncertainty Essentials

Wed, Sep 9

Conditional probability and base rates

Main: OpenIntro §§3.1–3.2

Probability, conditional percentages, independence, denominators, and reverse-conditional errors.

Mon, Sep 14

Distributions, randomness, and simulation

Main: OpenIntro §4.1; Handout H2

Distributions, the normal model, repeated samples, and simulation as a bridge to statistical inference; the sampling simulation run in R.

Unit 5 — Sampling, Polling, and Sampling Variation

Wed, Sep 16
Quiz 1
11:00–11:20

Quiz 1 and sampling in practice

Main: OpenIntro §1.3

Probability and nonprobability samples; coverage, nonresponse, and self-selection.

Mon, Sep 21

Sampling distributions, standard error, and the CLT

Main: OpenIntro §5.1; Handout H2

Statistics across samples; standard error; sample size; the Central Limit Theorem and its limits.

Unit 6 — Estimation and Testing

Wed, Sep 23

Confidence intervals and margins of error

Main: OpenIntro §§5.2 and 7.1

The interval structure built once, in full, for a proportion: estimate, standard error, margin of error, confidence level and its interpretation; the *t*-based interval for a mean as the same structure with a different standard error.

Mon, Sep 28

Hypothesis tests

Main: OpenIntro §5.3; Handout H2

Null hypotheses, *p*-values, decision errors, statistical significance versus substantive importance; a test computed and read in R.

Unit 7 — Comparing Groups and Categorical Evidence

Wed, Sep 30

Comparing means and proportions

Main: OpenIntro §§6.2 and 7.3

Comparing means across independent groups, developed in full; comparing proportions as the parallel categorical case; effect size, uncertainty, and research design; a group comparison shown in R.

Mon,Oct 5 – Fri,Oct 9 *Fall Break — no class.*

Mon, Oct 12 *Columbus Day/Indigenous People's Day — no class.*

Wed, Oct 14
Quiz 2
11:00–11:20

Quiz 2 and two-way tables, association, and confounding

Main: OpenIntro §§2.2.2 and 6.4; Agresti ch. 10 (recommended)

One two-way table carried through the whole session: joint, marginal, and conditional percentages; independence; confounding and Simpson's paradox as the culmination.

Unit 8 — Finding and Preparing Public Data

Mon, Oct 19

Inside two public data sources

Main: Handout H3

Two sources examined from the inside: what each holds, how the files and documentation are organized, how to read the metadata, and how to cite them; the two compared; what makes a question feasible. Other credible sources are mapped in H3, and any documented public dataset is welcome. **Project Guide (PG) posted on Canvas.**

Wed, Oct 21

From download to prepared data in R

Main: Handout H3

The full path shown once, with a predownloaded CSV and a supplied script: importing with `read.csv()`, inspecting variables, finding missing values, defining the analysis sample; students modify one filter and one recode themselves; keeping the raw file unchanged.

Unit 9 — The Regression Idea

Mon, Oct 26

From questions to evidence: scatterplots and correlation

Main: OpenIntro §8.1; Handout H4

Matching tables, comparisons, and plots to the question and variable types; scatterplots; correlation — strength, direction, and limits.

Wed, Oct 28

The least-squares line

Main: OpenIntro §8.2; Handout H4

Intercept, slope, fitted values, and residuals; the least-squares criterion and where the line comes from; interpreting the slope in substantive words.

Mon, Nov 2

Midterm examination

The 80-minute midterm covers Units 1–9 (all material through October 28). Calculator required; one handwritten cheat sheet permitted.

PART II — Applied Regression and the Data Analysis Project

Unit 10 — Simple Regression

Wed, Nov 4

Proposal due Fri, Nov 6

Estimating and interpreting the simple regression model

Main: OpenIntro §8.2; handout H4; Hanck et al. ch. 4

The model behind the line: how least squares chooses the intercept and slope; prediction and extrapolation; the same model estimated with `lm()` and its output translated into a substantive claim.

Mon, Nov 9

Uncertainty and fit

Main: OpenIntro §§8.2 and 8.4; Handout H4; Hanck et al. chs. 4–5

The standard error of the slope, confidence intervals, hypothesis tests, and R^2: what each can and cannot say; statistical versus substantive importance; the key numbers — coefficient, standard error, interval, and R^2 — read off R output in class.

Wed, Nov 11 *Veterans Day — no class.*

Unit 11 — Regression with Control Variables

Mon, Nov 16

Why add another variable?

Main: OpenIntro §9.1; Handout H4; Hanck et al. ch. 6

Confounding and omitted-variable logic; the multiple regression model and its ceteris-paribus interpretation; why adding controls can strengthen, weaken, reverse, or leave a relationship largely unchanged; a two-model comparison estimated in R.

Wed, Nov 18

Quiz 3
11:00–11:20

Quiz 3 and reading the two-model table

Main: OpenIntro §9.1; Handout H4

Presenting models side by side; reading a regression table column by column; regression figures; the same table built in R.

Unit 12 — Checking and Communicating Results

Mon, Nov 23

Residual warnings and the limits of a claim

Main: OpenIntro §9.3 ; Handout H4

Reading residual patterns as warning signs — curvature, unequal spread — and what they mean for your claims; association, prediction, and causal limits; residual plots produced and read in R.

Wed, Nov 25

Quiz 4
11:00–11:20

Quiz 4 and communicating regression results

Main: Handout H4; Project Guide (PG)

Integrating the regression table, figures, uncertainty, substantive interpretation, and limitations into one coherent empirical argument — applied directly to project data.

Thu–Fri, Nov 26–27 *Thanksgiving Break — no class.*

Unit 13 — Project Presentations

Mon, Nov 30

Final presentations I

Main: Student presentations

Wed, Dec 2

Final presentations II and course synthesis

Main: Student presentations

Remaining presentations; the semester workflow reviewed; where QPA II extends the toolkit.

Final Examination Period

Dec 7–11

Monday–Tuesday, December 7–8: Reading/Hurricane Makeup Days. Wednesday–Friday, December 9–11: no final examination. ***The data analysis paper is due Friday, December 11, at 11:59 p.m. on Canvas.***

Paper due
Fri, Dec 11