# POL 245: Visualizing Data

## Summary 2017

| | |
|---:|:---|
| Course Head: | Kosuke Imai |
| | kimai@princeton.edu |
| Instructor: | Will Lowe |
| | wlowe@princeton.edu |
| Preceptors: | Erik Wang |
| | haixiaow@princeton.edu |
| | Soichiro Yamauchi |
| | soichiro@princeton.edu |
| QuantLab: | Alex Tarr |
| | atarr@princeton.edu |

In this course, we consider ways to illustrate compelling stories hidden in a blizzard of data. Data visualization – equal parts art, programming, and statistical reasoning – is a critical tool for anyone doing analysis. In recent years, data analysis skills have become essential for those pursuing careers in policy advocacy and evaluation, business consulting and management, or academic research in the fields of education, health, medicine, and social science. This course introduces students to the powerful R programming language and the basics of creating data-analytic graphics in R. From there, we use real datasets to explore topics ranging from network data (like social interactions on Facebook or trade between counties) to geographical data (like county-level election returns in the US or the spatial distribution of insurgent attacks in Afghanistan). No prior background in statistics or programming is required or expected.

## Logistics

*The schedule during the first week deviates from this, details are below in the detailed course outline at the end of the syllabus.*

Google calendar for the course: http://goo.gl/D3pKXL

**Lectures.** Monday and Wednesday, 1:30pm–2:30pm, Louis A. Simpson A71. Lecture slides will appear on Blackboard immediately *after* the lecture. Students are expected to take notes during the lecture.

**Precepts.** Tuesday and Thursday, 1:30pm–2:50pm, Frist Campus Center 330, 309, and 209. Bring your personal laptop to precepts.

**QuantLabs.**  Monday (2:30-4:30pm), Tuesday (7-8:30pm), in the same room as your precepts. You will be working with tutors on review questions, practice exercises, and problem sets. Bring your laptop to the QuantLabs.

**Problem Set Help Sessions.**  Sunday (7-9pm) and Thursday (7-8:30pm) in Hargadon G001, G002, and G004

**Guest Lectures.**  Friday, 10:30–11:50am, Louis A. Simpson A71. These sessions occur during the second through final week of the course. They involve guest speakers from various industries where data visualization is used. Students should sign up for lunch with a specific speaker at the beginning of the course.

**Lunch with Guest Speaker.**  Friday, 12:00–1:30pm. Prospect House or at Mediterra, a downtown Princeton restaurant.

Students sign up to have lunch with one of the four guest speakers at the beginning of the course. During the selected week, students and the course team will meet with the guest speaker during a casual, catered lunch.

## Course Requirements

- **Class participation (15%):** Students should actively participate in all aspects of the course. Class participation will be judged based on questions asked/answered during the lectures and the precepts. Each portion is equally weighted.

- **Review Questions (15%):** During the QuantLab, students will work on the assigned portion of the textbook and electronically submit a small set of questions, using `Swirl`. Details on these assignments are announced at the QuantLab. *This is an individual assessment with limited collaboration.*

- **Problem sets (50%):** Each week will end with the posting of a problem set. These assignments can be retrieved by name using the `get_pset` function on the server. Electronic submission of your work must be uploaded to Blackboard by the beginning on Tuesday's precept. There is a short video on Blackboard showing the process.

  *This is an individual assessment with no collaboration.*

- **Final Project (20%):** This is a group data analysis project. Students will be assigned to groups. Analyzing a data set of their choice, students will write a report of no more than 1,000 words summarizing a compelling relationship or story they identified in the data. No more than 3 figures/tables can be used. Details regarding the final project will be announced later in the course. *This is a group assessment with collaboration allowed only within the assigned groups.*

  Final projects will be presented to the class at the end of the course.

## Collaboration Policy

The assignments in this course are designated as individual or group assessments. The degree of permissible collaboration depends on the kind of assignment:

- **Review Questions.** Students are encouraged to interact with each other, the instruction team, and QuantLab tutors in discussing their approaches and solutions. This includes conceptual discussion and actual computer code. *However, for all other assignments, this degree of collaboration is not appropriate!*

- **Problem Sets.** No collaboration is allowed. Students may ask clarifying questions regarding problem sets to the instruction team in person. This allows all students to benefit from clarifications equally. Clarifying questions about the problem sets may not be asked of QuantLab tutors, however.

- **Final Project.** Students may fully collaborate within their assigned groups, and may discuss their group's work with other students, the instruction team, and QuantLab tutors.

## Plagiarism Policy

Violations of the above collaboration policy will be treated as instances of plagiarism. This course will follow a modified version of the guidelines used for computer science classes here at Princeton. *Please take this guideline seriously.* In the past, plagiarism cases typically result in one-year suspension from Princeton.

Programming necessitates that you reach your own understanding of the problem and discover a path to its solution. DO NOT, UNDER ANY CIRCUMSTANCES, COPY ANOTHER PERSON'S CODE. Incorporating someone else's code into your program in any form is a violation of academic regulations. Abetting plagiarism or unauthorized collaboration by sharing your code is also prohibited. Sharing code in digital form is an especially egregious violation: do not e-mail your code to anyone.

Novices often have the misconception that copying and mechanically transforming a program (by rearranging independent code, renaming variables, or similar operations) makes it something different. Actually, identifying plagiarized source code is easier than you might think. For example, there exists computer software that can detect plagiarism.

This policy supplements the University's academic regulations, making explicit what constitutes a violation for this course. Princeton Rights, Rules, Responsibilities handbook asserts:

> The only adequate defense for a student accused of an academic violation is that the work in question does not, in fact, constitute a violation. Neither the defense that the student was ignorant of the regulations concerning academic violations nor the defense that the student was under pressure at the time the violation was committed is considered an adequate defense.

If you have any questions about these matters, please consult a member of the instruction team.

## Textbook

The course texbook is

Imai, Kosuke (2016). *Quantitative Social Science: An Introduction*. Princeton University Press.

## Statistical Software

In this course, we use the open-source statistical software R. R can be more powerful than other statistical software such as SPSS, STATA and SAS, but it can also be more difficult to learn. A variety of resources will be made available for POL 245 students in order to learn R as efficiently as possible. To help make using R easier, we'll be using RStudio — a user-interface that simplifies many common operations. You can find it here:

https://pol-rstudio.princeton.edu

## Get Help

Many students will find the materials in this course to be challenging. As such, students must seek immediate help when struggling with the course. There are several ways in which students can get in-person and online help.

### In-Person Help

- Office Hours: The preceptors will hold office hours. These take place at Monday 9-10:30am, Wednesday 2:30-4pm, and Thursday 3-4:30pm in Corwin 126. You will be able to ask any questions you might have about the course materials. You may also e-mail to set up an appointment outside of the office hours.

- Problem Set Help Sessions: Thursdays 7:00pm to 8:30pm and 7:00pm to 9:00pm on Sundays and in QuantLab, 2:30-4:30pm Monday and 7-8:30pm Tuesday. Tutors will not give you direct guidance on the actual problem set questions but will help you understand the concepts required for solving them.

## Introduction

During the first days of the course, you will be introduced to R statistical programming environment through the use of RStudio.

**T Jul 11**  1:30-2:30  Lecture: Introduction

Overview of the course.

2:30-3:50  Quantlab

Checking laptop setup and Swirl exercises.

Reading: ch. 1. Swirl: `INTRO1`, `INTRO2`

## Causality

We will learn how to infer causality from data. We learn the distinction between randomized experiments and observational studies. Our applications include the evaluation of strategies for increasing voter turnout and the effect of class size on educational achievement.

**W Jul 12**  1:30-2:50  Precept

Bias in turnout: `bias-in-turnout`

**R Jul 13**  1:30-2:30  Lecture: Causality

7:00-8:30  Quantlab

Reading: sec. 2.1–2.4. Swirl: `CAUSALITY1`

**F Jul 14**  10:30-11:50  Precept

Efficacy of small-class size in primary education: `small-class-size`

Problem set 1: Changing minds on gay marriage. `gay-marriage`

**S Jul 16**  7:00-9:00  Quantlab

Problem set help session.

---

**M Jul 17**  1:30-2:30  Lecture: Observational Studies

2:30-4:30  Quantlab

Reading: sec. 2.5–2.7. Swirl: `CAUSALITY2`

**T Jul 18**  1:30-2:50  Precept

Success of leader assassination as a natural experiment: `leader-assassination`

Problem Set 1 due.

7:00-8:30  Quantlab

Reading sec. 3.1–3.4. Swirl: `MEASUREMENT1`

## Measurement

We consider how to measure public opinion using sample surveys. We also learn about a measurement strategy regarding latent concepts like ideology. Our applications include surveys in Afghanistan and political polarization in US Congress.

**W Jul 19**   1:30-2:30   LECTURE: SURVEY SAMPLING

Surveys and sampling schemes.

**R Jul 20**   1:30-2:50   PRECEPT

Political efficacy in China and Mexico: `political-efficacy`

Problem set 2: Indiscriminate violence and insurgency: `indiscriminate-violence`

          7:00-8:30   QUANTLAB

**F Jul 21**   10:30-11:50   GUEST LECTURE

Dr. Y. Ghitza, Catalist Analytics

**S Jul 23**   7:00-9:00   QUANTLAB

Problem set help session.

---

**M Jul 24**   1:30-2:30   LECTURE: MEASUREMENT AND CLUSTERING

          2:30-4:30   QUANTLAB

Reading: sec. 3.5–3.8. Swirl: `MEASUREMENT2`

**T Jul 25**   1:30-2:50   PRECEPT

Voting in the United Nations General Assembly: `un-voting`

Due: Problem Set 2

          7:00-8:30   QUANTLAB

Reading: sec. 4.1. Swirl: `PREDICTION1`

**W Jul 26**   1:30-2:30   LECTURE: PREDICTION (AND LOOPS)

**R Jul 27**   1:30-2:50   PRECEPT

Prediction based on betting markets: `betting-markets`

Problem set 3: Oil, democracy, and development: `oil-democracy`

          7:00-8:30   QUANTLAB

**F Jul 28**   10:30-11:50   GUEST LECTURE

Ms. E. Safran, former White House Social and Behavioral Sciences Team

**S Jul 30**   7:00-9:00   QUANTLAB

Problem set help session.

**Prediction**

We learn about prediction starting with the application of US presidential election forecasting. Students will be introduced to linear regression and how it is related to causality.

| | | |
|---|---|---|
| **M Jul 31** | 1:30-2:30 | Lecture: Regression and causation |
| | 2:30-4:30 | Quantlab |
| | | Reading: sec. 4.2. Swirl: PREDICTION2 |
| **T Aug 1** | 1:30-2:50 | Precept |
| | | Prediction based on betting markets and linear models: betting-markets-with-lm |
| | | Due: Problem Set 3 |
| | 7:00-8:30 | Quantlab |
| | | Reading: sec. 4.3.1–4.3.3. Swirl PREDICTION3 |
| **W Aug 2** | 1:30-2:30 | Lecture: Regression and randomized experiments |
| **R Aug 3** | 1:30-2:50 | Precept |
| | | Elections and conditional cash transfer in Mexico: conditional-cash-transfers |
| | | Problem Set 4: Ideology of US Supreme Court justices: ideologies-of-justices |
| | 7:00-8:30 | Quantlab |
| **F Aug 4** | 10:30-11:50 | Guest Lecture |
| | | Ms. R. Lai, New York Times |
| **S Aug 6** | 7:00-9:00 | Quantlab |
| | | Problem set help session. |

| | | |
|---|---|---|
| **M Aug 7** | 1:30-2:30 | Lecture: Regression and Observational Studies |
| | 2:30-4:30 | Quantlab |
| | | Reading sec. 4.3.4 |
| **T Aug 8** | 1:30-2:50 | Precept |
| | | Government transfer and poverty reduction in Brazil: gov-transfer-brazil |
| | | Due: Problem Set 4 |
| | 7:00-8:30 | Quantlab |
| | | Reading sec. 5.1. Swirl: DISCOVERY1 |

**Discovery**

We cover how to analyze three different types of data; textual data, network data, and spatial data. Our applications include the prediction of disputed authorship of The Federalist Papers, the marriage network in Renaissance Florence, and the expansion of Wal-mart.

| | | |
|---|---|---|
| **W Aug 9** | 1:30-2:30 | LECTURE: TEXTUAL DATA |
| **R Aug 10** | 1:30-2:50 | PRECEPT |
| | | Analyzing the preambles of constitutions: `constitutions` |
| | 7:00-8:30 | QUANTLAB |
| **F Aug 11** | 10:30-11:50 | GUEST LECTURE |
| | | TBA |
| **S Aug 13** | 7:00-8:30 | QUANTLAB |
| | | Problem set help session. |

---

| | | |
|---|---|---|
| **M Aug 14** | 1:30-2:30 | LECTURE: NETWORK DATA |
| | 2:30-4:30 | QUANTLAB |
| | | Reading sec. 5.2. Swirl: `DISCOVERY2` |
| **T Aug 15** | 1:30-2:50 | PRECEPT |
| | | The international trade network: `trade-networks` |
| | 7:00-8:30 | QUANTLAB |
| | | Reading sec. 5.3. Swirl: `DISCOVERY3` |
| **W Aug 16** | 1:30-2:30 | LECTURE: SPATIAL DATA |
| **R Aug 17** | 1:30-2:50 | PRECEPT |
| | | Spatial mapping of US election results over time: `mapping-elections` |
| | 7:00-8:30 | QUANTLAB |
| **F Aug 18** | 10:30-11:50 | LECTURE |
| | | Wrapping up. |
| **S Aug 20** | 7:00-9:00 | QUANTLAB |

---

| | | |
|---|---|---|
| **T Aug 22** | 3:00-4:00 | FINAL PROJECT PRESENTATIONS |