Term Information

Effective Term	Autumn 2015
Previous Value	*Spring 2015*

Course Change Information

What change is being proposed? (If more than one, what changes are being proposed?)

To allow POLITSC 3780 to fulfill the Data Analysis requirement of the General Education curriculum.

What is the rationale for the proposed change(s)?

Political Science 3780 provides students with the basic concepts of statistices and probability, teaches them the methods necessary to analyze and critically evaluate statistical arguments, and fosters the importance of statistical ideas.

What are the programmatic implications of the proposed change(s)?

(e.g. program requirements to be added or removed, changes to be made in available resources, effect on other programs that use the course)?

none

Is approval of the requrest contingent upon the approval of other course or curricular program request? No

Is this a request to withdraw the course? No

General Information

Course Bulletin Listing/Subject Area	Political Science
Fiscal Unit/Academic Org	Political Science - D0755
College/Academic Group	Arts and Sciences
Level/Career	Undergraduate
Course Number/Catalog	3780
Course Title	Data Literacy and Data Visualization
Transcript Abbreviation	Data Lit & Vis
Course Description	Introduction to the tools of data analysis in political science, with an emphasis on data visualization.
Semester Credit Hours/Units	Fixed: 3

Offering Information

Length Of Course	14 Week, 7 Week, 4 Week (May Session), 12 Week (May + Summer)
Flexibly Scheduled Course	Never
Does any section of this course have a distance education component?	No
Grading Basis	Letter Grade
Repeatable	No
Course Components	Recitation, Lecture
Grade Roster Component	Recitation
Credit Available by Exam	No
Admission Condition Course	No
Off Campus	Never
Campus of Offering	Columbus

Prerequisites and Exclusions

Prerequisites/Corequisites

Exclusions

Cross-Listings

Cross-Listings

Subject/CIP Code

Subject/CIP Code	45.1001
Subsidy Level	Baccalaureate Course
Intended Rank	Freshman, Sophomore, Junior, Senior

Requirement/Elective Designation

Required for this unit's degrees, majors, and/or minors

General Education course:

Data Analysis

Previous Value

Required for this unit's degrees, majors, and/or minors

Course Details

Course goals or learning objectives/outcomes

- Develop skills in drawing conclusions and critically evaluating results based on data.
- Understand basic concepts of statistics and probability.
- Comprehend methods needed to analyze and critically evaluate statistical arguments.
- Recognize the importance of statistical ideas.

Content Topic List

- Locating and obtaining data
- Design of data visualizations
- R programming
- Creating visualizations (distributional, spatial, and temporal, etc.)
- The fundamentals of probability and uncertainty
- Monte Carlo simulations for understanding probability
- The bootstrap: generalizing from a sample to a larger population
- Research design: theories, hypotheses, and hypothesizing

Attachments

- Pols3780-syllabus.pdf: syllabus 3780

  (Syllabus. Owner: Smith,Charles William)

- 3780 rationale and objectives.pdf: rationale and objectives

  (Other Supporting Documentation. Owner: Smith,Charles William)

- Assessment-plan (2).pdf: GENED assessment

  (GEC Course Assessment Plan. Owner: Smith,Charles William)

- Stats concurrence.pdf: Concurrence from STATS

  (Concurrence. Owner: Smith,Charles William)

Comments

- Please see guidelines for GE Data Analysis pp. 43-45. Concurrence from Dept of Statistics is not uploaded. *(by Vankeerbergen,Bernadette Chantal on 11/05/2014 02:47 PM)*

Workflow Information

Status	User(s)	Date/Time	Step
Submitted	Smith,Charles William	10/28/2014 12:29 PM	Submitted for Approval
Approved	Herrmann,Richard Karl	10/28/2014 02:20 PM	Unit Approval
Approved	Haddad,Deborah Moore	10/28/2014 05:35 PM	College Approval
Revision Requested	Vankeerbergen,Bernadette Chantal	11/05/2014 02:47 PM	ASCCAO Approval
Submitted	Smith,Charles William	12/15/2014 02:53 PM	Submitted for Approval
Approved	Herrmann,Richard Karl	12/15/2014 03:35 PM	Unit Approval
Approved	Haddad,Deborah Moore	12/15/2014 05:11 PM	College Approval
Pending Approval	Nolen,Dawn Vankeerbergen,Bernadette Chantal Hanlin,Deborah Kay Jenkins,Mary Ellen Bigler Hogle,Danielle Nicole	12/15/2014 05:11 PM	ASCCAO Approval

Data Literacy and Data Visualization

Political Science 3780

Syllabus

Instructor Information
Bear F. Braumoeller
The Ohio State University
Department of Political Science
Office: Derby 2168
Office hours: Wednesday, 2:00–4:00 p.m., OBA
e-mail: braumoeller.1@osu.edu

Spring Semester, AY 2014-15
Syllabus version 1.0

Course description

Many, if not most, of the major debates in modern political science revolve around questions that can be addressed with data. The sources of voting behavior, the correlates of war, the determinants of development, political economy, psychology, institutions, and conflict—all are issues that are amenable to data-based analysis.

At the same time, the amount of available data and the number of publicly-available open-source tools for cleaning, transforming, analyzing and visualizing it have increased exponentially since the turn of the millennium. With a few clicks students can compare word frequencies in books over time or construct elaborate size-weighted wordclouds—tasks that would have taken scholars weeks if not months of effort in the past.

This course introduces students to those tools and the principles behind their use in the context of applications in political science. It marries the substance of political theory to the methodologies of data visualization and exploratory data analysis. It neither requires nor imparts any statistical background: it is designed to serve either as a standalone course or as a gateway to a more advanced data-analytics class.

Requirements

The format of the course is unusual: the lectures are online, and the Professor and the TA will meet weekly with the students in a small computer lab to work one-on-one with you through homework and exercises. Completing the exercises will require having listened to and viewed the lectures prior to class, but students may do so at whatever time is convenient. Questions about lecture material should be raised in the online Carmen forums, where they can be answered by the Professor and the TA.

The lectures are recorded in Quicktime format, which should be viewable on nearly any computer or mobile device. One advantage to the format is that, on most devices, lectures can be sped up to 1.4x to 1.5x; beyond about 2x, most of them become unintelligible. Students are advised to take advantage of this feature, as humans can generally understand speech at a higher rate than they can produce it. Be advised, though, that higher speeds generally require more focused attention, as important details are easier to miss. It's also possible to rewind and slow the recording back down if a particular section moves quickly or is difficult to understand.

Students will attend all recitation sections. There will not be a traditional midterm or final exam. Rather, weekly short assignments will make up 60% of the grade, and the remaining 40% will come from a final project in which the student finds a dataset in his or her area of interest that is not already used in the course, analyzes it to assess the structure of the data, and works through the most appropriate, succinct, and informative summaries and visualizations. Students will be given the last 2-3

weeks of recitation sections to work in-class on final projects, with the Professor and the TA present to assist.

The final project is designed to be the foundation of an independent research project that will fulfill the requirements for a senior thesis. Students who are eligible to graduate with research distinction who are interested in writing a senior thesis should discuss doing so with their advisor very soon, as applications are due early in the fall semester. Funding for thesis research is available from the University. Applications for the Undergraduate Research Scholarship (URS) and the International Research Grant for undergraduates in the Arts and Sciences are available at `http://aschonors.osu.edu/opportunities/scholarships/undergrad`.

Again, students who wish to compete for funding for a senior project should speak to their advisors soon. The deadline for applications and project advisor recommendations is September 16, and selections will be made by mid-November. Applicants will compete for approximately 50 scholarships awarded in amounts ranging from $500 to $6,000. University regulations require that the funds be used to meet tuition and fee expenses, with any remaining amount available directly to recipients (provided they have not already exceeded the maximum allowable amount for financial aid).

Books

Three books are recommended for the class. There is no coursepack. The books are:

Tufte, Edward. *The Visual Display of Quantitative Information.* (Graphics Pr)

Yau, Nathan. *Visualize This: The FlowingData Guide to Design, Visualization, and Statistics.* (Wiley)

Hrishi, Mittal. *R Graph Cookbook* (Kindle edition; Packt Publishing)

Raykov, Tenko, and George A. Marcoulides. *Basic Statistics: An Introduction with R.* (Rowman & Littlefield)

GE Goals and Expected Learning Outcomes: Data Analysis

Goals

Students develop skills in drawing conclusions and critically evaluating results based on data.

Expected Learning Outcomes

Data Analysis—Students understand basic concepts of statistics and probability, comprehend methods needed to analyze and critically evaluate statistical arguments, and recognize the importance of statistical ideas.

Academic Misconduct

It is the responsibility of the Committee on Academic Misconduct to investigate or establish procedures for the investigation of all reported cases of student academic misconduct. The term "academic misconduct" includes all forms of student academic misconduct wherever committed; illustrated by, but not limited to, cases of plagiarism and dishonest practices in connection with examinations. Instructors shall report all instances of alleged academic misconduct to the committee (Faculty Rule 3335-5-487). For additional information, see the Code of Student Conduct (http://studentaffairs.osu.edu/resource_csc.asp).

All students believe that they know how not to plagiarize. Many of them are wrong. Every year, many of them find that out the hard way. Don't be one of them.

The short version is that passing off another person's work or ideas as your own is plagiarism. That includes the unacknowledged word-for-word use or paraphrasing of another person's work or ideas. It is not enough, for example, simply to copy and paste a passage and then cite the source at the end. If the passage is taken word-for-word, it must be in quotes as well to indicate that fact.

There is an excellent video at http://hdl.handle.net/1811/46848, if you have any doubts. You should be crystal clear, as the University's policies exist to ensure fairness, and violators of University regulations on academic integrity will be dealt with severely.

Disability Services

Students with disabilities that have been certified by the Office for Disability Services will be appropriately accommodated, and should inform the instructor as soon as possible of their needs. The Office for Disability Services is located in 150 Pomerene Hall, 1760 Neil Avenue; telephone 292-3307, TDD 292-0901; http://www.ods.ohio-state.edu/.

Lecture Schedule

Date	To view before class
August 26	Read syllabus
September 2	Labor Day (no classes)
September 9	Lectures 1 and 2
September 16	Lectures 3 and 4
September 23	Lectures 5 and 6
September 30	Lectures 7 and 8
October 7	Lectures 9 and 10
October 14	Lectures 11 and 12
October 21	Lectures 13 and 14
October 28	Lectures 15 and 16
November 4	Lectures 17 and 18
November 11	Veterans' Day (no classes)
November 18	Lectures 19 and 20
November 25	Lectures 21 and 22
December 2	Lectures 23 and 24

Lectures

Lecture 1: Introduction. Nature of the course, its relationship to other courses and students' plans of study; how and why data visualization can be useful.

Lecture 2: Data are All Around Us! Terminology, sources, and advice on what to do if you can't find a preexisting dataset. The difference between observational and experimental studies. (Yau, pp. 21–43)

Lecture 3: Data Tools, part I. Open-source solutions that require little or no additional information or effort to produce compelling results.

Lecture 4: The Good. Exemplary representations of data from our field and others, with a discussion of what makes them so compelling. (Tufte, ch. 1; Yau, pp. 2–12)

Lecture 5: The Bad and The Ugly. Ideas that seemed good at the time, why they really aren't, how we can fix them, and what general principles they imply. (Tufte, ch. 2; Yau, 13–20)

Lecture 6: Data Tools, part II. These versatile tools allow you to upload, explore, and visualize your own data. (Yau, pp. 54–62)

Lecture 7: Truth. Data visualization is inherently a reduction of information. The main imperative, in reducing information, is integrity: representing the data as honestly as possible. We look at some good cases and some not-so-good cases. (Tufte, ch. 3)

Lecture 8: Introduction to R. A powerful (and free) general statistical package, `R` is capable of producing a wide range of graphs. In these lectures we demonstrate how to download and install the package, add libraries, import data, and create visualizations. (Yau, pp. 71–75.)

Lecture 9: Beauty. The difference between a fairly good data visualization and an amazing one often lies in the application of a few straightforward graphical principles. (Tufte, chs. 4–6)

Lecture 10: R, part II. Exploration of basic plots.

Lecture 11: Time. Different ways of examining variables over time. (Yau, ch. 4)

Lecture 12: R, part III. Advanced topics in `R` graphics.

Lecture 13: Space. Less common, but more eye-catching, are ways of looking for patterns across space, especially in maps. (Yau, pp. 80–89)

Lecture 14: R, part IV. Practicum: tailoring `R` graphics to specific problems.

Lecture 15: Space and Time. Representing relationships and trends over both space and time. (Yau, ch. 5)

Lecture 16: Comparisons. Learn to compare the values of multiple variables.

Lecture 17: Research Design 101. Theories, hypotheses, research designs, tests, and conclusions.

Lecture 18: Notions of Probability. The axioms of probability and basic probability calculations. (R & M, ch. 4)

Lecture 19: Notions of Probability. Random variables and probability calculations using random variables. (R & M, ch. 5)

Lecture 20: Notions of Probability. Expected values. (R & M, ch. 6)

Lecture 21: Statistical Inference. Moving from a sample to a population; bias and variance.

Lecture 22: Statistical Inference. Understanding the margin of error and confidence. (R & M, ch. 9)

Lecture 23: Statistical Inference. The logic of statistical testing. (R & M, ch. 11)

Lecture 24: Statistical Inference. The misuse of statistics.

Rationale

Many undergraduate data analytics courses focus on the mechanics of data analysis without first conveying fundamental data literacy. The result is rather like what would happen if we taught Hemingway, in English, to people who didn't speak English: their efforts to pick up the language along the way dramatically compound the challenges of absorbing the subject matter. This course conveys fundamental data literacy in the context of visualization, which provides obvious and engaging motivation. It then introduces students to the R statistical language, one of the most powerful tools in the world for statistical analysis. Finally, it shows them how to understand basic concepts in statistics and probability – uncertainty, generalizing from a sample to a population, etc. – using simulations and resampling. It is meant to provide both a foundation and a motivation for further quantitative training and research.

Learning objectives/outcomes

Students develop skills in finding and generating datasets, manipulating, summarizing, and visualizing data, critically evaluating arguments using data, and measuring the uncertainty of their conclusions.

The course will fulfill the General Education (GE) requirement in Data Analysis by helping students develop skills in drawing conclusions and critically evaluating arguments based on data. It will introduce students to basic concepts in statistics and probability, including sampling, data distributions, and the Central Limit Theorem, and it will teach students how to use iterated simulation and resampling (i.e., Monte Carlo simulation and bootstrapping) to obtain estimates of unknown probabilistic outcomes and to assign measures of accuracy to sample estimates. It will tie these elements together with the logic of research design in order to give students the ability to evaluate statistical arguments, and it will show them examples of how to do so using data on such topics as American partisanship and elections, drone strikes, and international conflict.

THE OHIO STATE UNIVERSITY

College of Arts and Sciences
Department of Political Science
2140 Derby Hall
154 N Oval Mall
Columbus, OH 43210
614-292-2880 Phone

GE Assessment Plan
Political Science 3780
Professor Braumoeller

Overview. In order to assess aggregate student achievement of GE expected learning outcomes over time, the faculty will examine both raw (un-curved) assignment scores on the weekly and Final Projects and the number of questions asked on each assignment in the Carmen forums. The former is a direct indicator of performance; the latter is an indirect indicator that measures students' difficulty in completing the assignments unaided.

Description of the specific methods the faculty will use to demonstrate that the aggregate of his/her students are achieving the goals and expected learning outcomes of this GE category.

Students will be given a rubric for each weekly assignment and the Final Project, and individual students' performance will be evaluated vis-à-vis the criteria laid out in the rubric. An example (Assignment 3) is attached, along with a rubric (Figure 1).

A separate discussion forum will be set up for each assignment so that students who are encountering difficulty can post questions. The number of questions is an indirect indicator of students' ability to achieve the learning outcomes: although no questions at all may indicate that the assignment is insufficiently challenging, a large number is indicative that students are struggling to complete it. An example (Figure 2) is attached. In that Figure we can see that, while most Assignments occasioned little discussion, Assignment 4 was characterized by mass confusion.

Finally, one very useful indirect indicator of overall learning is item 7 on the standardized evaluation of instruction (SEI), which asks whether the students agree that they "learned greatly from the instructor."

Explanation of the level of student achievement expected

The assignments build on one another: a skill that is introduced in one assignment will be honed and perfected over the following assignments. For this reason, there is no direct correspondence between a grade on any one assignment and the degree to which a particular skill has been learned. The best overall indicator of a student's level of skill mastery is the Final Project (attached), which counts for 40% of the final grade and requires students to utilize all of the skills that they have learned in the class. Success will be defined as 75% or more of the students receiving grades in the "Excellent" category (80-100) on the un-curved Final Project grade.

THE OHIO STATE UNIVERSITY

College of Arts and Sciences

Department of Political Science

2140 Derby Hall
154 N Oval Mall
Columbus, OH 43210

614-292-2880 Phone

For the discussion forum indicator, experience suggests that, even on a successful assignment, it is reasonable to expect a number of posts less than or equal to roughly 20% of the class size, indicating that up to 10% of the students have asked questions—generally points of clarification or idiosyncratic issues—and received answers. Success will be defined as a number of posts equal to or less than 20% of the class size.

For item 7 on the SEI, an average score of 4.0 or better will be taken as an indication that the students themselves perceived a successful learning outcome.

Description of follow-up/feedback process

Once the faculty has gathered data on student achievement, s/he will use this information to determine, in consultation with senior colleagues and other specialists at the University if necessary, which lectures need to be improved, which assignments need to be changed (or replaced), and which texts need to be added or removed.

The information—scores, rubrics, discussion forum totals, and SEIs—will be archived on Carmen and in the Faculty Center.

THE OHIO STATE UNIVERSITY

College of Arts and Sciences

Department of Political Science

2140 Derby Hall
154 N Oval Mall
Columbus, OH 43210

614-292-2880 Phone

Maps and Data	Excellent 4 points	Good 2 points	Poor 0 points
World maps	Two separate GunnMap world maps, one reflecting the percentage of the population in poverty in each country, the other reflecting percentage obese. Color schemes and balance chosen so that full range of variation is clear. Either JPG or PNG format.	One GunnMap world map, correct, as described above; or two maps with incorrect data or imbalanced/unuseful color scheme.	No maps submitted.
Network data and graph	Correct network data in csv format; map of network that usefully displays network structure.	Incorrect or incomplete network data, or map of network that has had no layout applied (a circular blob, in Gephi, or a rectangle, in Cytoscape) and doesn't reflect network structure.	Missing data and network map.
Essay	**Excellent 2 points**	**Good 1 point**	**Poor 0 points**
Response essay	One paragraph correctly describing the structure of the *Francs-tireurs partisans* network and offering a plausible explanation of why it might look the way it does. Word, text, or PDF format.	An incorrect description of the structure of the network, or no attempt to explain why it might look the way it does.	No response essay submitted.
Overall Score	**Excellent 8 or more**	**Good 5 or more**	**Poor 0 or more**

Figure 1. Rubric for Assignment 3

THE OHIO STATE UNIVERSITY

College of Arts and Sciences

Department of Political Science

2140 Derby Hall
154 N Oval Mall
Columbus, OH 43210

614-292-2880 Phone

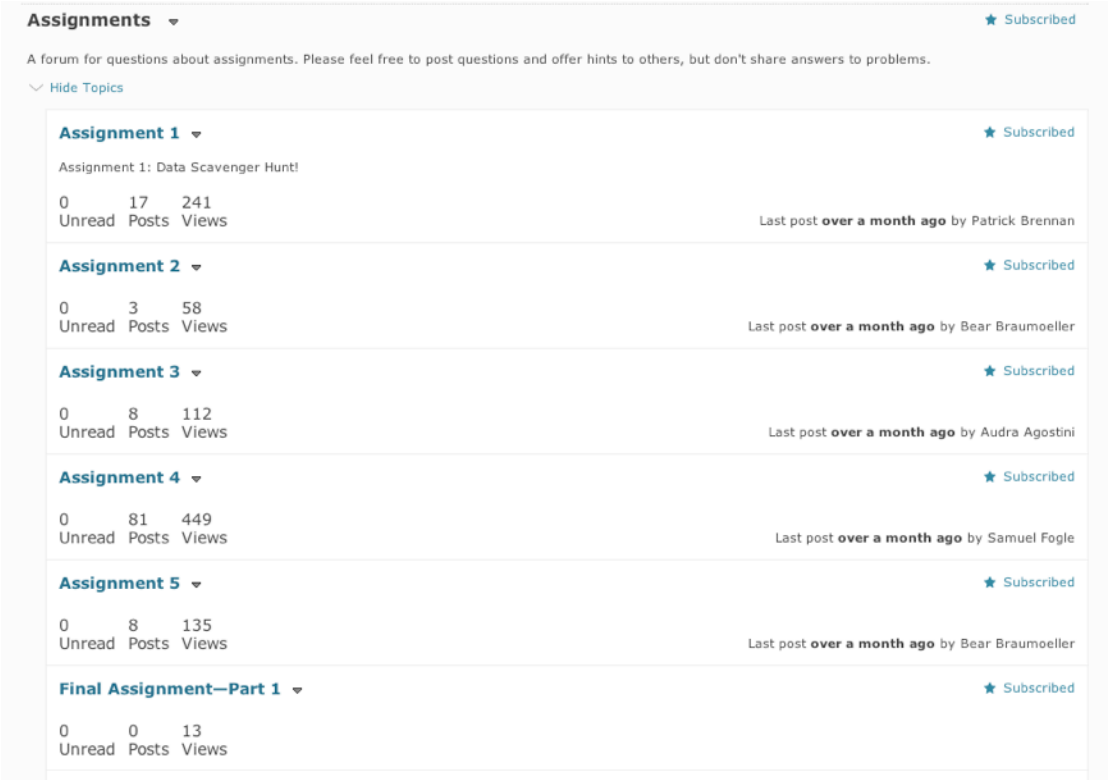

Figure 2. Discussion forum post totals.

THE OHIO STATE UNIVERSITY

College of Arts and Sciences

Department of Political Science

2140 Derby Hall
154 N Oval Mall
Columbus, OH 43210

614-292-2880 Phone

Setup	Excellent 5 points	Good 3 points	Poor 0 points
Question	A succint (ideally, one sentence) and logically coherent research question.	A vague or unfocused research question that could be answered once specified more clearly.	A research question that is entirely lacking in coherence, or no research question.
How these data help	A description of the data used to answer the question, and an explanation of why these data and this data analysis will help us to answer the question.	An incomplete description of the data used to answer the question, and/or a superficial explanation of why these data and procedures will help us to answer the question.	Little or nothing by way of description of the data or explanation of their utility.
Process	**Excellent 10 points**	**Good 7 points**	**Poor 0 points**
Process of answering question	Detailed description of the process by which you obtained and analyzed data, including all details that a reader would need to reproduce the process perfectly.	Brief description of the process by which you obtained and analyzed data, with enough detail that an informed reader could probably reproduce your results with a bit of guesswork.	No description, or a description so vague as to be useless.
Answer	**Excellent 20 points**	**Good 15 points**	**Poor 0 points**
Your answer	A compelling data analysis that gives a clear, convincing, thorough answer to the question utilizing visualizations that score high on the dimensions of truth and beauty; also, a candid discussion of any remaining doubts and how they might be resolved by future research.	An adequate data analysis that gives a plausible answer to the question using visualizations, and a cursory discussion of any remaining doubts.	An analysis that really doesn't answer the question at all, or no analysis.
Overall Score	**Excellent 35 or more**	**Good 25 or more**	**Poor 0 or more**

Figure 3. Rubric for Final Project

THE OHIO STATE UNIVERSITY

College of Arts and Sciences

Department of Political Science

2140 Derby Hall
154 N Oval Mall
Columbus, OH 43210

614-292-2880 Phone

Maps and Data	Excellent 4 points	Good 2 points	Poor 0 points
World maps	Two separate GunnMap world maps, one reflecting the percentage of the population in poverty in each country, the other reflecting percentage obese. Color schemes and balance chosen so that full range of variation is clear. Either JPG or PNG format.	One GunnMap world map, correct, as described above; or two maps with incorrect data or imbalanced/unuseful color scheme.	No maps submitted.
Network data and graph	Correct network data in csv format; map of network that usefully displays network structure.	Incorrect or incomplete network data, or map of network that has had no layout applied (a circular blob, in Gephi, or a rectangle, in Cytoscape) and doesn't reflect network structure.	Missing data and network map.
Essay	**Excellent 2 points**	**Good 1 point**	**Poor 0 points**
Response essay	One paragraph correctly describing the structure of the *Francs-tireurs partisans* network and offering a plausible explanation of why it might look the way it does. Word, text, or PDF format.	An incorrect description of the structure of the network, or no attempt to explain why it might look the way it does.	No response essay submitted.
Overall Score	**Excellent 8 or more**	**Good 5 or more**	**Poor 0 or more**

Figure 1. Rubric for Assignment 3

THE OHIO STATE UNIVERSITY

College of Arts and Sciences

Department of Political Science

2140 Derby Hall
154 N Oval Mall
Columbus, OH 43210

614-292-2880 Phone

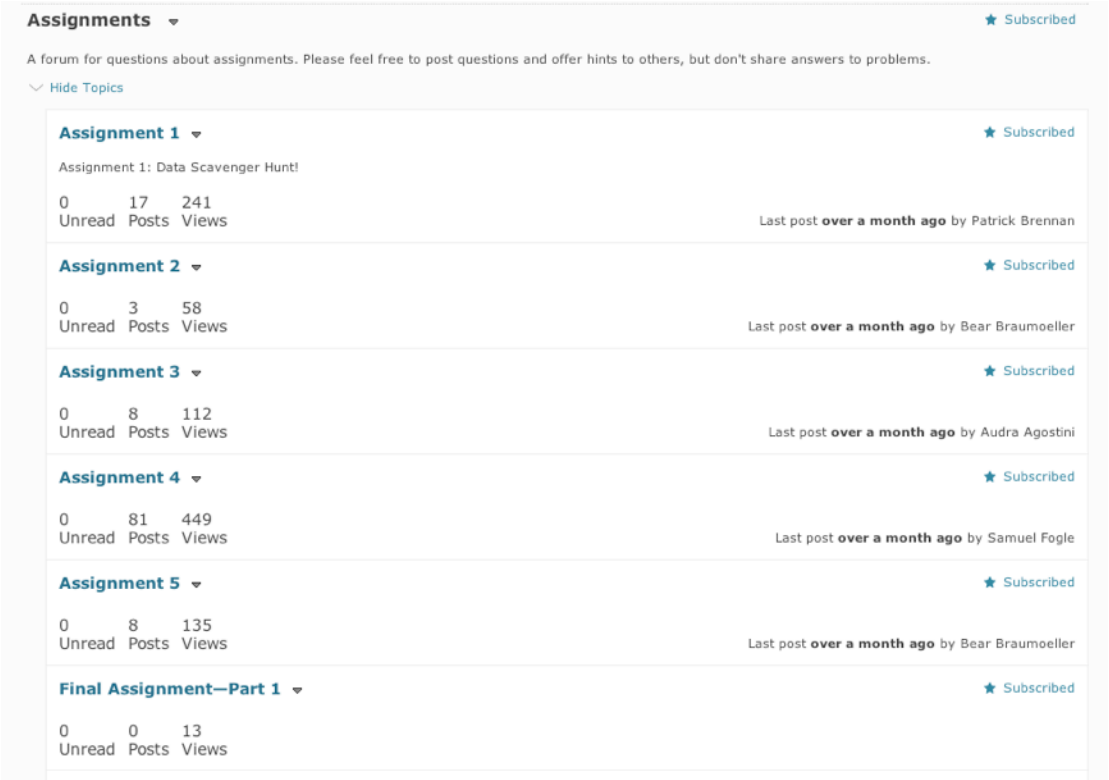

Figure 2. Discussion forum post totals.

THE OHIO STATE UNIVERSITY

College of Arts and Sciences

Department of Political Science

2140 Derby Hall
154 N Oval Mall
Columbus, OH 43210

614-292-2880 Phone

Setup	Excellent 5 points	Good 3 points	Poor 0 points
Question	A succint (ideally, one sentence) and logically coherent research question.	A vague or unfocused research question that could be answered once specified more clearly.	A research question that is entirely lacking in coherence, or no research question.
How these data help	A description of the data used to answer the question, and an explanation of why these data and this data analysis will help us to answer the question.	An incomplete description of the data used to answer the question, and/or a superficial explanation of why these data and procedures will help us to answer the question.	Little or nothing by way of description of the data or explanation of their utility.
Process	**Excellent 10 points**	**Good 7 points**	**Poor 0 points**
Process of answering question	Detailed description of the process by which you obtained and analyzed data, including all details that a reader would need to reproduce the process perfectly.	Brief description of the process by which you obtained and analyzed data, with enough detail that an informed reader could probably reproduce your results with a bit of guesswork.	No description, or a description so vague as to be useless.
Answer	**Excellent 20 points**	**Good 15 points**	**Poor 0 points**
Your answer	A compelling data analysis that gives a clear, convincing, thorough answer to the question utilizing visualizations that score high on the dimensions of truth and beauty; also, a candid discussion of any remaining doubts and how they might be resolved by future research.	An adequate data analysis that gives a plausible answer to the question using visualizations, and a cursory discussion of any remaining doubts.	An analysis that really doesn't answer the question at all, or no analysis.
Overall Score	**Excellent 35 or more**	**Good 25 or more**	**Poor 0 or more**

Figure 3. Rubric for Final Project

Assignment 3

Overview. This assignment consists of two parts. The first involves setting up a web scraper. It may take a couple of attempts and will probably take a while to run, so you might consider setting it up first and then working on the second part of the assignment while it's running.

Part I: Stuffed and Starved

Noted author Raj Patel is coming to Ohio State to give a talk. His book, *Stuffed and Starved: The Hidden Battle for the World Food System,* explores the reasons behind the simultaneous epidemics of starvation and obesity worldwide. Word of your data-visualization skills has spread among your friends, and you've been contacted by the editor of the *Lantern* and asked to produce two world maps for the newspaper—one exploring the distribution of poverty in the world and the other exploring the distribution of obesity.

Go to the CIA World Factbook website and click on "View Text/Low Bandwidth Version" for the uglier (but much more scraper-friendly) version of the site. Use Outwit Hub to collect data on (a) the adult prevalence rate of obesity and (b) the percentage of the population below the poverty line for as many countries as you can.[1]

Using GunnMap, create two graphs representing "The Stuffed World" (obesity rate) and "The Starved World" (percentage below the poverty line). Use whatever color schemes you'd like, but be sure that the balance is set so that readers can see the full spectrum of variation. Save the maps as `stuffed.jpg` and `starved.jpg`.

Part II: The Structure of Insurgency Networks

The *Francs-tireurs* were irregular riflemen who engaged in guerrilla warfare and what we would now call covert operations. Their history dates back at least to the Franco-Prussian War, though they became most prominent during the French Resistance in World War II.

[1] There will be a fair bit of missing data, both because some governments don't report these figures and because the CIA World Factbook collects data on territories as well as autonomous countries. Don't worry too much about it; GunnMap handles missing data pretty well.

Your assignment is to analyze the network structure of the *Francs-tireurs partisans* (FTP), the military arm of the French Communist Party, which engaged the Nazis as part of the French Resistance following the German invasion of the Soviet Union.

The data for connections among 175 members of the FTP have been collected by Alexander Gutfraind at Cornell University. To download and analyze them, visit Mr. Gutfraind's website at http://www.cam.cornell.edu/~gfriend/research.php and find the link that says "Network data on the underground network Francs-tireurs et Partisans." Either download the data or copy-and-paste them into a text file.

In order to prepare the data to be read into a network-graphing program (Gephi or Cytoscape), you will need to do the following in a text editor:

- Remove all the comments at the top (the lines that start with "#").
- Do a global search and replace within the text file to replace all spaces with commas.
- Save the file, with a .csv suffix (for example, "FTP.csv").

You will then need to open the CSV file in a spreadsheet program like Excel and do the following:

- Delete the third column—the one that only contains 1s.
- Insert a blank row of cells at the top of the spreadsheet (in Excel, you do this by clicking on row 1, going to the Insert menu, and selecting Rows).
- At the top of the first column, enter the word "Source". At the top of the second column, enter the word "Target".
- Save as a CSV file (for example, "FTP2.csv").

You now have an *edge list,* or a list of all of the connections between nodes in the network. Load this file into either Gephi or Cytoscape and explore a few different layout options for the data. When you find one that gives you a good sense of what the network looks like (hint: it's not just a blob), save a copy of the network graph.

In a separate Word or text file, answer the following question: Why do you think the *Francs-tireurs partisans* network looks the way it does? (1 paragraph)

Upload to Assignment 3 Dropbox a compressed folder containing (a) your two maps from Part I, (b) your CSV file and network graph from Part II, and (c) your answer to the question in Part II.

Final Project, Part 1

Your final project for this class involves asking and answering an interesting question, using data visualization. In this, the first part of that assignment, you will ask the question.

Choose a topic that you've studied (in a political science class or elsewhere), or a topic that you'd like to know more about, and **propose a question** that

1. is interesting
2. has not already been answered (or, has been answered, but could be answered differently)
3. can be answered with data

Explain, in one paragraph each, why it interests you, how you could answer it, and where you could find data that would answer it (include specific sources and datasets).

A question should be neither too broad ("What causes the rise of empires?") nor too narrow ("How popular was the President last year?") Shoot for a question that can be explored using available data in under 10 pages. The best way to do this is to focus on a manageable part of a broad question that interests you: rather than trying to ascertain whether gun control increases or decreases crime in general, for example, you might explore specific gun-related laws that have been enacted in some states but not others—perhaps find two states that are as similar as possible, except that one passed the law and one didn't, and see what their crime rates look like before and after the law's passage.

Final Project, Part 2

Your final project for this class involves asking and answering an interesting question, using data visualization. In this, the second part of that assignment, you will ask and answer the question. You do not need to ask the same question, exactly, but if your question is substantially different, make sure that you've run it past me or Ms. Bradshaw for feedback.

1. State the question and explain why you find it interesting.
2. Explain why the data you examine will help you to answer the question.
3. Explain, step by step, how you've gone about answering it. Describe the data and where you obtained the them, what (if anything) you did to reformat or transform them, how you analyzed them, and what they told you. Include visualizations. (This should be most of the project.)
4. What do you now know that you didn't know before? Does the answer raise further questions that might be worth investigating? If so, describe them briefly.

We anticipate papers in the 5–7pp. range, 1$^1/_2$-space, though succinct writers may take less space and those with more complex problems or answers may take more. You may use any programs or websites you'd like to format and analyze the data; `R` should be very useful for this assignment, but it is not required.

Due to the University's strict timeline for final grades, no extensions can be offered except in case of genuine emergency. We look forward to receiving your best effort by 5:00 p.m. on December 4.

Department of Statistics

404 Cockins Hall
1958 Neil Avenue
Columbus, OH 43210-1247

Phone (614) 292-2866
Fax (614) 292-2096
Web www.stat.osu.edu

December 12, 2014

Bear F. Braumoeller
Associate Professor and Director of Graduate Studies
Department of Political Science
2168 Derby Hall
154 North Oval Mall
Columbus, OH 43210

Dear Professor Braumoeller;

On behalf of the Department of Statistics, I am happy to give concurrence to Political Science 3780, Data Literacy and Data Visualization, to count toward the GE requirement in Data Analysis. The course meets the requirements of the GE in Data Analysis and looks to be very interesting. I hope your students enjoy the course and you enjoy teaching it.

Sincerely,

William Notz

William Notz
Vice Chair for Administration and Undergraduate Studies
Department of Statistics
The Ohio State University