Term Information

Effective Term	Spring 2017

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	3780H
Course Title	Data Literacy and Data Visualization
Transcript Abbreviation	Data Lit & Vis
Course Description	Most social science debates can be addressed with data, and sources of data are growing exponentially. This course introduces students to tools of data analysis and principles behind their use in the context of social-science applications.
Semester Credit Hours/Units	Fixed: 3

Offering Information

Length Of Course	14 Week, 12 Week, 8 Week, 7 Week, 6 Week, 4 Week
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

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

- syllabus3780.pdf: syllabus non-honors 3780

  (Syllabus. Owner: Smith,Charles William)
- 3780 statement of difference.pdf: Statement of Difference

  (Statement of Qualitative Difference. Owner: Smith,Charles William)
- 3780H Assessment plan.pdf: GEC assessment plan

  (GEC Course Assessment Plan. Owner: Smith,Charles William)
- 3780H GE rationale and learning outcomes.pdf: GE rationale and learning outcomes

  (Other Supporting Documentation. Owner: Smith,Charles William)
- SyllabusHonors3780.pdf: syllabus Honors 3780

  (Syllabus. Owner: Smith,Charles William)
- Assignment F1.pdf: course assignment F1

  (Other Supporting Documentation. Owner: Smith,Charles William)
- Assignment F2.pdf: course assignment F2

  (Other Supporting Documentation. Owner: Smith,Charles William)
- Assignment 3.pdf: course assignment 3

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

Comments

- The Assignments included (F1, F2, 3) address recurring typo on pp 35-38 of course proposal. *(by Smith,Charles William on 11/02/2016 04:56 PM)*
- See e-mail from M Toohey to J Mitzen and C Smith *(by Vankeerbergen,Bernadette Chantal on 10/31/2016 04:41 PM)*

Workflow Information

Status	User(s)	Date/Time	Step
Submitted	Smith,Charles William	10/18/2016 11:32 AM	Submitted for Approval
Approved	Herrmann,Richard Karl	10/18/2016 01:04 PM	Unit Approval
Approved	Haddad,Deborah Moore	10/18/2016 03:28 PM	College Approval
Revision Requested	Vankeerbergen,Bernadette Chantal	10/31/2016 04:42 PM	ASCCAO Approval
Submitted	Smith,Charles William	11/02/2016 04:57 PM	Submitted for Approval
Approved	Herrmann,Richard Karl	11/02/2016 05:01 PM	Unit Approval
Approved	Haddad,Deborah Moore	11/02/2016 05:29 PM	College Approval
Pending Approval	Nolen,Dawn Vankeerbergen,Bernadette Chantal Hanlin,Deborah Kay Jenkins,Mary Ellen Bigler Hogle,Danielle Nicole	11/02/2016 05:29 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

January 25, 2016

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 to help students better understand the social world. Moreover, the course introduces students to basic concepts in statistics and probability so that they can understand, not just the most plausible answer to their questions, but the degree of uncertainty surrounding that answer. 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 in that the lectures are all online. In addition to the lectures, there will be a fifty-minute weekly recitation section in which the Professor and the TA will meet with the students in a 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 required for the class and one is recommended. There is no coursepack. The books are:

> Yau, Nathan. *Visualize This: The FlowingData Guide to Design, Visualization, and Statistics.* (Wiley)
>
> Teetor, Paul. *R Cookbook.* (O'Reilly Cookbooks)
>
> Carsey, Thomas M. and Jeffrey J. Harden. *Monte Carlo Simulation and Resampling Methods for Social Science.* (Sage)
>
> Chang, Winston. *R Graphics Cookbook.* (O'Reilly Cookbooks; recommended)

GE Goals and Expected Learning Outcomes

Goals

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

Achieving Expected Learning Outcomes

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.

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.

Assessment of Expected Learning Outcomes

The effectiveness of this course in achieving the expected learning outcomes outlined above will be determined in three ways.

1. A problem set will be assigned in class every week in weeks 2–13 to help students internalize the material delivered in the online lectures. The problem sets will be graded, students' mastery of the relevant skills will be assessed, and they will be informed of problem areas, if any.

2. During the last two weeks of class, students will work on an independent project of their choosing, one that involves answering questions using the skills they have acquired. This final assignment will count for 40% of the course grade.

3. Students will be urged to fill in the online Student Evaluation of Instruction (SEI) reports to assess the amount that they have learned in this course relative to others.

Grading

Appropriately for a course of this nature, grading takes place via visualization. Course grades are assigned on a curve, with the exact cutoffs depending on the distribution of student performance. Typically, grades in the top quintile receive As, those in the second quintile receive low As or high Bs, those in the third quintile receive Bs and low Bs, and those in the fourth and fifth quintiles receive Cs, Ds, or Es.

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
Week 1	Read syllabus
Week 2	Lectures 1 and 2
Week 3	Lectures 3 and 4
Week 4	Lectures 5 and 6
Week 5	Lectures 7 and 8
Week 6	Lectures 9 and 10
Week 7	Lectures 11 and 12
Week 8	Lectures 13 and 14
Week 9	Lectures 15 and 16
Week 10	Lectures 17 and 18
Week 11	Lectures 19 and 20
Week 12	Lectures 21 and 22
Week 13	Lectures 23 and 24
Week 14	Lectures 25 and 26
Week 15	Lectures 27 and 28

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. (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. (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. (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.

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, and import data. (Yau, pp. 71–75; Teetor, chs. 1–4.)

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. (Teetor, ch. 5.)

Lecture 10: R, part II. Data structures and data transformations. (Teetor, chs. 6–7.)

Lecture 11: APIs. How to connect to a vast array of public data sources.

Lecture 12: Plots in R. Scatterplots, line graphs, boxplots, dot charts, and more. (Teetor, chs. 8–10.)

Lecture 13: Time. Different ways of examining variables over time. (Yau, ch. 4)

Lecture 14: Space. Less common, but more eye-catching, are ways of looking for patterns across space, especially in maps. (Yau, pp. 80–89)

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

Lecture 16: Basic Probability I. Set theory, probability, conditional probability, how probabilities cumulate. (Carsey and Harden, ch. 2.1–2.2.)

Lecture 17: Basic Probability II. Random variables and probability distributions. (Carsey and Harden, ch. 2.3–2.6.)

Lecture 18: Simulated Worlds. Simulating probability distributions, creating distribution functions, systematic and stochastic components. (Carsey and Harden, ch. 4.)

Lecture 19: Samples and Populations I. What's in a label? The power of permutations. (Carsey and Harden, ch. 8.1–8.2.)

Lecture 20: Samples and Populations II. Resampling, or, pulling yourself up by the bootstrap. (Carsey and Harden, ch. 8.4.)

Lecture 21: Research Design. Theories, hypotheses, research designs, tests, and conclusions. (Carsey and Harden, ch. 7.)

Lecture 22: Workshop—Election 2012. How do political scientists predict election outcomes with such shocking accuracy?

Lecture 23: Workshop—Partisanship. Is American politics more partisan than it was 50 years ago? If so, why?

Lecture 24: Workshop—Drone Strikes. The pros and cons of warfare via unmanned aerial vehicle.

Lecture 25: Workshop—The End of War? Is war becoming more common, or is it actually on the decline?

Lecture 26: Workshop—The Austerity Debate. Is it better to spend one's way out of recession or cut back further?

Lecture 27: Workshop—Gaza. Exploring patterns of attacks and reciprocity.

Lecture 28: Conclusion: How to Lie Without Statistics. What are the limits of inference via data visualization? What more can data tell us that we can't access visually or intuitively? Worse, how can apparently obvious visual patterns mislead us? In short, why do people ever take statistics courses?

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

November 1, 2016

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 to help students better understand the social world. Moreover, the course introduces students to basic concepts in statistics and probability so that they can understand, not just the most plausible answer to their questions, but the degree of uncertainty surrounding that answer. 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 in that the lectures are all online. In addition to the lectures, there will be a fifty-minute weekly recitation section in which the Professor will meet with the students in a 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.

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 two projects in which the student (a) produces a useful visualization for an existing political science paper and (b) 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 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

Five books are required for the class and one is recommended. There is no coursepack. The books are:

> Yau, Nathan. *Visualize This: The FlowingData Guide to Design, Visualization, and Statistics.* (Wiley)
>
> Tufte, Edward. *The Visual Display of Quantitative Information.* (Graphics Press)
>
> Teetor, Paul. *R Cookbook.* (O'Reilly Cookbooks)
>
> Carsey, Thomas M. and Jeffrey J. Harden. *Monte Carlo Simulation and Resampling Methods for Social Science.* (Sage)
>
> Zieffler, Andrew S., Jeffrey R. Harring, and Jeffrey D. Long. *Comparing Groups: Randomization and Bootstrap Methods Using R.* (Wiley)

Chang, Winston. *R Graphics Cookbook.* (O'Reilly Cookbooks; recommended)

GE Goals and Expected Learning Outcomes

Goals

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

Achieving Expected Learning Outcomes

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.

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.

Assessment of Expected Learning Outcomes

The effectiveness of this course in achieving the expected learning outcomes outlined above will be determined in three ways.

1. A problem set will be assigned in class every week in weeks 2–13 to help students internalize the material delivered in the online lectures. The problem sets will be graded, students' mastery of the relevant skills will be assessed, and they will be informed of problem areas, if any.

2. In week 11, students will be required to hand in a short paper in which they obtain data from a published study in political science and generate a useful, original visualization designed to convey some of the findings of the original study. This assignment will count for 15% of the course grade.

3. During the last two weeks of class, students will work on an independent project of their choosing, one that involves answering questions using the skills

they have acquired. This final assignment will count for 25% of the course grade.

4. Students will be urged to fill in the online Student Evaluation of Instruction (SEI) reports to assess the amount that they have learned in this course relative to others.

Grading

Per Honors guidelines, straight grades will be assigned for this course: the final grade will not be curved.

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 098 Baker

Hall, 113 W. 12th Ave.; telephone 614-292-3307, TDD 292-0901; http://www.ods.ohio-state.edu/.

Lecture Schedule

Date	To view before class
Week 1	Read syllabus
Week 2	Lectures 1 and 2
Week 3	Lectures 3–5
Week 4	Lectures 6 and 7
Week 5	Lectures 8 and 9
Week 6	Lectures 10 and 11
Week 7	Lectures 12 and 13
Week 8	Lectures 14 and 15
Week 9	Lectures 16–18
Week 10	Lectures 19–21
Week 11	Lectures 22–24
Week 12	Lectures 25–27
Week 13	Lectures 29 and 30
Week 14	Lectures 31 and 32
Week 15	Lectures 33, 34, and 28

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. (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, and import data. (Yau, pp. 71–75; Teetor, chs. 1–4; Zieffler, Harring, and Long, ch. 1.)

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. Data structures and data transformations. (Teetor, chs. 5–7; Zieffler, Harring, and Long, ch. 2.)

Lecture 11: APIs. How to connect to a vast array of public data sources.

Lecture 12: Plots in R. Scatterplots, line graphs, boxplots, dot charts, and more. (Teetor, chs. 8–10; Zieffler, Harring, and Long, ch. 4.1–4.4.)

Lecture 13: Time. Different ways of examining variables over time. (Yau, ch. 4)

Lecture 14: Space. Less common, but more eye-catching, are ways of looking for patterns across space, especially in maps. (Yau, pp. 80–89)

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

Lecture 16: Basic Probability I. Set theory, probability, conditional probability, how probabilities cumulate. (Carsey and Harden, ch. 2.1–2.2.)

Lecture 17: Basic Probability II. Random variables and probability distributions. (Carsey and Harden, ch. 2.3–2.6.)

Lecture 18: Simulated Worlds. Simulating probability distributions, creating distribution functions, systematic and stochastic components. (Carsey and Harden, ch. 4.)

Lecture 19: Samples and Populations I. What's in a label? The power of permutations. (Zieffler, Harring, and Long, ch. 6; Carsey and Harden, ch. 8.1–8.2.)

Lecture 20: Samples and Populations II. Resampling, or, pulling yourself up by the bootstrap. (Zieffler, Harring, and Long, ch. 7; Carsey and Harden, ch. 8.4.)

Lecture 21: Research Design. Theories, hypotheses, research designs, tests, and conclusions. (Carsey and Harden, ch. 7.)

Lecture 22: Workshop—Election 2012. How do political scientists predict election outcomes with such shocking accuracy?

Lecture 23: Workshop—Partisanship. Is American politics more partisan than it was 50 years ago? If so, why?

Lecture 24: Workshop—Drone Strikes. The pros and cons of warfare via unmanned aerial vehicle.

Lecture 25: Workshop—The End of War? Is war becoming more common, or is it actually on the decline?

Lecture 26: Workshop—The Austerity Debate. Is it better to spend one's way out of recession or cut back further?

Lecture 27: Workshop—Gaza. Exploring patterns of attacks and reciprocity.

Lecture 28: Conclusion: How to Lie Without Statistics. What are the limits of inference via data visualization? What more can data tell us that we can't access visually or intuitively? Worse, how can apparently obvious visual patterns mislead us? In short, why do people ever take statistics courses?

Lecture 29: Regression, loess, and splines. Visualizing trends in data or relationships among variables is often done via simple regression, but

other, more flexible methods are available as well.

Lecture 30: Visualizing Uncertainty. The degree of uncertainty is often crucial information to convey in a visualization, but it is too often omitted. This lecture explores different ways in which it can be conveyed.

Lecture 31: Experiments and Natural Experiments. The logic behind experiments, why the evidence they produce is so compelling, and how to find natural experiments.

Lecture 32: Difference-in-Difference and Regression Discontinuity. Types of natural experiment, the logic behind each, and the mechanics of visualizing the results.

Lecture 33: The Challenges of Thick-Tailed Distributions. Income, city size, and battle deaths in war all follow thick-tailed distributions that are challenging to visualize.

Lecture 34: Separating Signal from Noise. Is there really a change in the average value of a variable over time, or are we simply witnessing normal random variation in its value? We explore a straightforward permutation-based test for answering this question and methods for visualizing the result.

Data Literacy and Data Visualization

Political Science 3780

Statement of Qualitative Difference

September 15, 2016

The ASCC Honors Panel expects that honors courses will differ from non-honors courses in a variety of ways and so requires that the proposer include a statement that addresses the following items (with particular attention to the differences between the two versions of the course, if a non-honors version exists):

1 How the specific goals of the course will be achieved

The goals of the course are that students develop skills in drawing conclusions and critically evaluating results based on data. The expected learning outcomes are that they will understand basic concepts of statistics and probability, comprehend methods needed to analyze and critically evaluate statistical arguments, and recognize the importance of statistical ideas.

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.

These goals and objectives will be achieved via lectures, readings, recitation sections, and assignments. The format of the course is unusual in that the lectures are all online. **Non-Honors students will have readings drawn from a total of 3 books; Honors students will also have more in-depth readings drawn from an additional 2 books.** In addition to the lectures and readings, there will be a fifty-minute weekly recitation section in which the Professor will meet with the students in a computer lab to work one-on-one with them 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 can also be raised in the online Carmen forums, where they can be answered by the Professor.

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 two projects in which the student (a) produces a useful visualization for an existing political science paper and (b) 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 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. **Whereas students in the non-Honors version of the course are asked for a data analysis that is not to exceed 10pp. in length, Honors students will be required to go into more depth about their argument, the relevant literature, the data required to test the argument, and the visualizations themselves, resulting in a paper of approximately 25pp. in length.**

2 The exposure to the basic material in the course, and ways in which added breadth and depth of material will be included.

Exposure to basic material will be through lectures, readings, recitation sections and assignments, as described above. **Added depth and breadth for Honors students will be achieved via a 21% increase in the number of lectures and a corresponding increase in readings, with more in-depth readings assigned from more advanced texts.**

3 The exposure to, and use of, methodology and research techniques, and especially the ways in which the course will provide exposure to the nature of scholarship in the field.

The additional lecture materials and readings for Honors students will focus entirely on more advanced methodologies and research techniques, and the additional assignments are designed to give Honors students a greater opportunity to explore political science scholarship.

4 Amount and quality of work expected from students on papers, examination(s), and projects; and the method of grading that work.

Honors students will be held to a higher standard on assignments and projects than non-Honors students, and their grades will not be curved. While the number of weekly assignments will not change (since the number of weeks does not change), the weekly assignments will be more challenging. The additional project and final project, described above, will also require more and better work from Honors students.

5 The amount and kind of student/faculty contact, including how the course will offer a significant level of interaction and engagement between faculty and students, and how such engagement will be achieved.

Political Science 3780 already offers a tremendous amount of student-faculty contact: because the lectures are available online, the Professor does nothing *but* interact with the students as they work on weekly assignments. This is one of the best features of the course. **In the Honors version of the class, no TA will be present, so every student will have unfettered access to the Professor during all recitation sections.**

6 How an environment will be fostered that facilitates intellectual exchange among students (if applicable).

In both the Honors and the non-Honors versions of the course, students are given the opportunity to complete weekly assignments in groups of no more than 3. In past years I have found that this structure greatly enhances intellectual exchange among students. There will be no difference between Honors and non-Honors courses in this regard.

7 Ways that creative thinking will be an essential aspect of the course requirements.

Creative thinking is essential to completing the weekly assignments and the final assignments in the non-Honors version of the course. Indeed, creativity is at the

core of exploratory data analysis. **The Honors version will require more work and, therefore, greater creativity, and it will provide students with more of a foundation for that work.**

8 How the course will embrace, as appropriate, interdisciplinary work and study.

As a GE course, 3780 draws students from many different majors. Accordingly, the examples from the course are drawn from fields as diverse as political science, economics, medicine, journalism, and business. The techniques taught for visualization are all inherently cross-disciplinary in application, and students are free to devise a final project in whatever field they'd like.

9 Evidence of a pedagogical process that will demand a high level of intellectual output.

11 weekly assignments, 1 final paper (**longer and more detailed for Honors students**), and **1 additional in-depth assignment for Honors students** require constant engagement with the material and a high level of mastery.

Political Science 3780H

GE rationale & expected learning outcomes: Data Analysis

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

Expected Learning Outcomes: 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.

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.

(7) GE assessment plan

The effectiveness of this course in achieving the expected learning outcomes outlined above will be determined in three ways.

1. A problem set will be assigned in class every week in weeks 2-13 to help students internalize the material delivered in the online lectures. The problem sets will be graded, students' mastery of the relevant skills will be assessed, and they will be informed of problem areas, if any.

2. During the last two weeks of class, students will work on an independent project of their choosing, one that involves answering questions using the skills they have acquired. This final assignment will count for 40% of the course grade.

3. Students will be urged to fill in the online Student Evaluation of Instruction (SEI) reports to assess the amount that they have learned in this course relative to others.

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 3780H
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

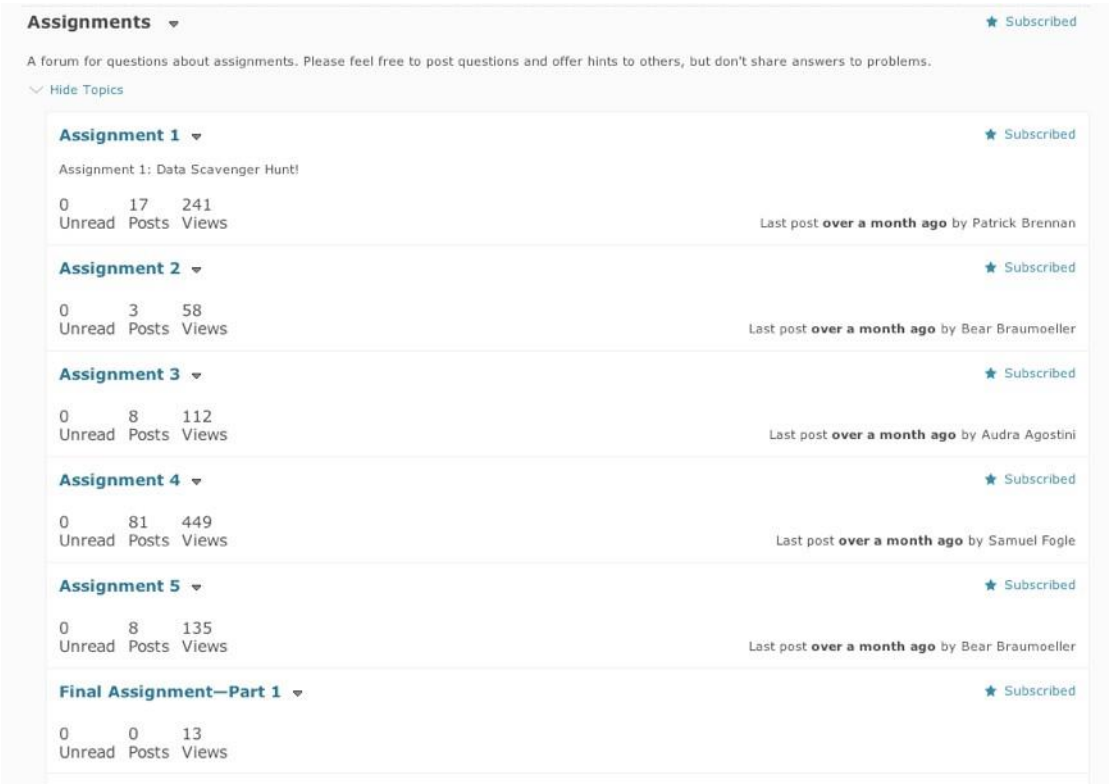

Figure 2. Discussion forum post totals.

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

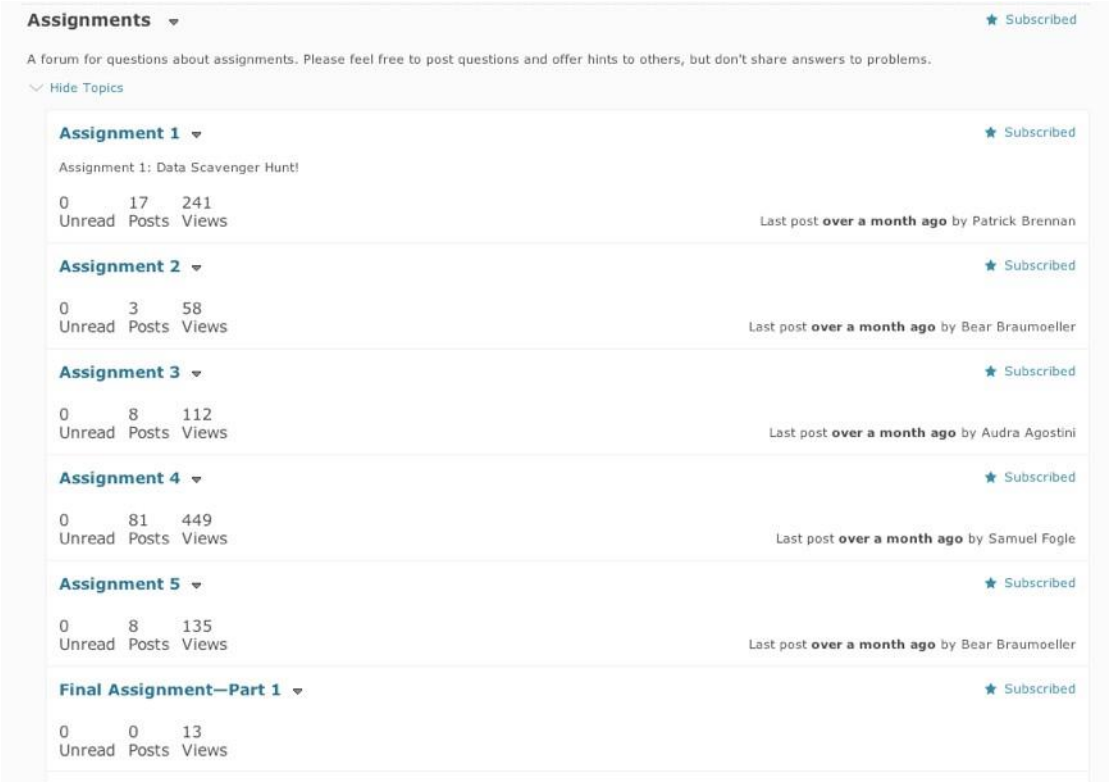

Figure 2. Discussion forum post totals.

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 êrst 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 êrst 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.[κ]

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 riëemen 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.

κThere will be a fair bit of missing data, both because some governments don't report these êgures 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 ênd 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 êle.

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 êle to replace all spaces with commas.
- Save the êle, with a .csv suffix (for example, "FTP.csv").

You will then need to open the CSV êle 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 êrst column, enter the word "Source". At the top of the second column, enter the word "Target".
- Save as a CSV êle (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 êle into either Gephi or Cytoscape and explore a few different layout options for the data. When you ênd 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 êle, 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 êle and network graph from Part II, and (c) your answer to the question in Part II.

Final Project, Part 1

Your ênal project for this class involves asking and answering an interesting question, using data visualization. In this, the êrst 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 ênd data that would answer it (include speciêc 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 speciêc gun-related laws that have been enacted in some states but not others—perhaps ênd 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 ênal 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 ênd 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 brieëy.

We anticipate papers in the 5–7pp. range, $1\frac{1}{2}$-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 ênal 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.

Final Project, Part 1

Political Science 3780

Due: Tuesday, March 10, at 5:00 p.m.
in Assignment F1 Carmen dropbox

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. 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 in under 10 pages, using available data. *Do not simply reproduce an existing analysis.* 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

Political Science 3780

Due: Monday, April 27, at 5:00 p.m.
in Final Assignment Carmen dropbox

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. Yi 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 April 27.

Assignment 3

Political Science 3780

Due: Friday, September 13, at 5:00 p.m.
in Assignment 3 Carmen dropbox

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://gutfraind.com, look under "Research," 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, PDF, 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.