

Syllabus

**Time, Place, and (Social) Data: Advanced Issues in Large-Scale Analysis and Visualization**

Instructor:  
Jacob Groshek, Ph.D.

# 1. Practical Information

<b>Course Name</b>	Time, Place, and (Social) Data: Advanced Issues in Large-Scale Analysis and Visualization
<b>Course Code</b>	EM847
<b>Credits</b>	4
<b>Term</b>	Fall 2015
<b>Instructor</b>	Dr. Jacob Groshek 704 Commonwealth Ave. 302D <a href="mailto:jgroshek@bu.edu">jgroshek@bu.edu</a> 617-353-6421
<b>Office Hours</b>	6:00pm – 9:00pm Tuesdays Also by appointment
<b>Location</b>	871 Commonwealth Ave. (CGS) 121
<b>Timetable</b>	Section A1 – 6:00pm until 9:00pm Tuesdays
<b>Course Website</b>	<a href="https://www.dropbox.com/home/EM888#">https://www.dropbox.com/home/EM888#</a>
<b>Compulsory Literature</b>	Reinard, J. C. (2006). Communication Research Statistics. Sage: Thousand Oaks, CA. ISBN: 9780761929871  Stock & Watson (2007). Introduction to Econometrics. Pearson: New York, NY. ISBN-13: 978-0-321-27887-6
<b>Additional Literature</b>	<i>Additional readings are available online, on Dropbox or via the BU library.</i>

## 2. Course Overview

Wk.	Date	Topic	Literature / Assignments	Outcome
1	9-08	Introduction and Analytic Background	CRS: Ch1, 2, 3, and 4 Exercise 1 and 2	Competence in summary statistics of sample data
2	9-15	Inferential Statistics in Large Datasets	CRS: Ch7, 8, 9, and 10 Exercise 3	Competence in comparing means and categories with sample data
3	9-22	Unit Project	Student-provided readings Unit Project 1	Successfully complete <b>original statistical routines through means and categories with</b> analyses and interpretations
4	9-29	Correlation and Regressions: Ordinary, Nonlinear, and Binary	CRS: Ch5, 13, 14; Stock & Watson, Ch9 (also consider Ch4, 5, 6) Exercise 4	Competence in correlating and regressing variables with sample data in SPSS and Stata
5	10-06	Unit Project	Student-provided readings Unit Project 2	Successfully complete <b>original statistical routines through nonlinear regression</b> analyses and interpretations
6	<b>10-13 (no class)</b>	Pretend	it is	Monday?
7	10-20	Scaling Factors and Panel Data	CRS: Ch15; Stock & Watson, Ch8 Exercise 5	Competence in creating <b>reliable scales</b> and regressing variables with sample <b>panel</b> data

8	10-27	Workshop	Individual meetings	Revisions based on individual feedback
9	11-03	Unit Project	Student-provided data Unit Project 3	Successfully complete <b>original factoring and panel data</b> analyses and interpretations
10	11-10	Prediction and Causality in Time Series	Stock & Watson, Ch12, 13, and 14 Exercise 6	Competence in regressing variables with sample time series data in Stata
11	11-17	Unit Project	Student-provided readings Unit Project 4	Successfully complete <b>original time series</b> analyses and interpretations of secondary data
12	11-24	Data Mining and Predicting	Franch, 2013; Groshek, 2012; Meier, 2011; Bollen et al., 2011; Exercise 7	Construct social data time-series models that analyze and predict relationships
13	12-01	Unit Project	Student-provided readings Unit Project 5	Successfully <b>regress original social data</b> and <b>create original visualizations</b>
14	12-08	Final Project Presentations	Student-provided readings	Develop models to explore, see, and answer questions—and successfully present research to an audience

### **3. Course Introduction**

This course provides a specialized emphasis on data processing and predictive modeling through time-series and panel regression modeling. In so doing, it trains students in advanced social-scientific methods for large-scale data analysis and visualization. This course also incorporates approaches that integrate the analysis and graphing of social data using both time and spatial models. In addition, the use of advanced software in emerging media research is developed.

Though this course is positioned to harness the potential of social data, one of its main themes is to equip students to synthesize social data into more traditional and historical datasets. In this way, students leave prepared to do analyses of an historically informed kind, which is used as a jumping off point for models that predict events in the future. Within this temporal framework, students also learn the importance of place and develop models to contribute meaningfully to the discipline.

#### **Instructor Information**

Jacob Groshek holds a PhD from Indiana University (2008) in Mass Communications, has a Master's degree in Mass Communication from Marquette University (2003), and earned a BS in Technology Education from the University of Wisconsin. He is an internationally recognized scholar in studying emerging media technologies as their content and use may relate to political change at the macro (national) and micro (individual) levels. He has over 25 peer-reviewed research publications and has spoken at more than 50 international conferences, including research talks and workshops as well as national-level policy statements in venues such as the US State Department in Washington DC, the London School of Economics, City University of Hong Kong, and the Institute for Advanced Study in Toulouse, France. He has received multiple awards for his research and sits on editorial boards of leading journals in the field.

He currently is an Assistant Professor of Emerging Media at Boston University, where he spearheads innovative research methods and developing advanced models of mining and interpreting social media data. One of his more notable contributions to the field has been his commitment to advancing the use of time-series and panel data analyses to communication research. Professor Groshek has travelled the world to land at BU, having previously held similar appointments at the University of Melbourne (Australia), the Erasmus University Rotterdam (Netherlands), and Iowa State University. He is an active member of the Association of Education in Journalism and Mass Communication (AEJMC), where he headed the Communication Technology division, the International Communication Association (ICA), the Association of Internet Researchers (AoIR), and the International Association of Media and Communication Research (IAMCR).

A native of Wisconsin, he is now a husband and father of two (ages 8 and 4). With his family, he enjoys exploring New England--especially in the fall--and coaching youth sports when not reviewing papers or grading. Previously he was a home brewer and finds statistics nearly as important to civilization as good beer, even though he typically now drinks crisp, mineral-ly whites and dry red wines. He also enjoys writing autobiographies in the third person and has been known to be just a little bit sarcastic. He admires hard working students that do readings and ask questions, especially those that force him to learn more.

## 4. Course Objectives and Outcomes

*Students have the knowledge and understanding of:*

1. How to use analyze large datasets;
2. how to engage with various software systems, namely SPSS, Stata, Excel, StatTransfer, R, and TCAT;
3. how and when to use these software platforms to perform advanced statistical and visual modeling in large datasets;
4. how to work with panel and time-series data in order to a) analyze, b) explain, and c) predict specific events;
5. what the benefits and limitations of these approaches are, and how those can be minimized or negotiated.

*Students have the ability to:*

1. Analyze and interpret large-scale data with time-series and panel approaches;
2. create meaningful visualizations of that data using leading software;
3. locate and collect secondary data for original analyses;
4. develop time-series models that explain and predict outcomes from (social) data;
5. carry out original research that uses time and space to examine and explain relationships that otherwise may be overlooked or misinterpreted;
6. produce an original manuscript that is suitable for submission and presentation to an academic conference;
7. make a contribution to the body of knowledge in a given cognate area;
8. use analyses and visualizations to further the development of communication theory.

## 5. Organization & Working Method

### General

It is necessary to attend class sessions because without them there is no course. Also, because the class meets only once per week missing one class session is seriously problematic.

It is therefore compulsory to attend all class meetings, arrive on time, and to participate actively in the discussions and other activities. This obligation includes the preparation and submission of all assignments. Attendance will be taken, and students are not permitted absences without penalty. Any work missed will need to be submitted in due course with an appropriate late penalty assessed. Travel plans will not excuse anyone from the deadline for submitting assignment(s).

Students are advised to prepare for each class by studying the readings. In certain weeks, handouts of example articles will be distributed. Classes are designed to (a) give the opportunity for questions on the literature, (b) help students in working on the assignments, and (c) provide specific feedback on assignments.

## Use of Blackboard

Blackboard is horrible. We will never use it. Ever.

## **6. Assessment and Grading**

Students will complete 8 weekly exercises and 5 unit projects, all of which are to be completed individually. Students are, however, welcomed to collaborate in solving problems, but each must work independently with unique data for each exercise.

The deadline for all exercises is Monday morning by noon before the next class session. Assignments will not be accepted beyond the deadline, and submissions will not be accepted in improper format (i.e., no hardcopies).

### **Description of the exercises**

The exercises are all a combination of empirical and practical research considerations using sample data where students must seek to answer substantive research questions. Each exercise asks students to apply the discussed readings and carry out some form of statistical analysis and/or visualization toward the end of advancing knowledge. Each exercise will be laid out in full during the weekly class session.

### **Description of the unit projects**

The unit projects are where students move out of sample data analyses and start doing original research, either with secondary data or with original data collection. Each unit project requires students to apply the discussed readings and carry out some form of statistical analysis and/or visualization based on not just one week of content but a combination of topics over several weeks and to synthesize topics within the units. Each unit project will be laid out in full during the appropriate class session.

### **Criteria and grading**

#### *Criteria used to evaluate all assignments*

The assignments will be evaluated based on accuracy, appropriateness, clarity, and quality of their work. Although students are not evaluated explicitly for their English writing abilities, they are expected to check upon their language before handing in assignments. This means: (1) during the writing process students should consult a dictionary when they are not sure if a word or phrase is correct and (2) after the writing process they should use spell-check before handing in their work.

#### *Criteria for grading of participation*

Each student's active participation is vital to the success of this course. As such, the participation grade is measured with a combination of contributions to in-class discussions in relation to the readings and the lectures. Attendance, punctuality, and effective cooperation within class sessions are considered part of the participation grade. There will be an opportunity for self-evaluation as necessary before the end of the term.

### *Grading*

Every assignment will be evaluated on a 100 scale. Your final score will be calculated based on the above percentages, which will then be translated to your final letter grade using the following formula:

93-100	A	73-76.99	C
90-92.99	A-	70-72.99	C-
87-89.99	B+	67-69.99	D+
83-86.99	B	63-66.99	D
80-82.99	B-	60-62.99	D-
77-79.99	C+	0-59.99	F

### *Weight of assessments*

Table 1: Overview of different elements for grading

Types of assignment	Points possible	Team/Individual	Percent of final grade
Final Project	120	Team (optional)	40.0%
Final Project Presentation	40	Individual	13.3%
Unit Projects (x 5)	15 (x5 = 75)	Team (optional)	25.0%
Exercises (x 8)	5 (x8 = 40)	Individual	13.3%
Participation	25	Individual	8.4%
Total	300		100%

### **Intellectual Integrity**

In accordance with the high standards of excellence set forth by, and for, all members of the Boston University community, the College of Communication finds it imperative that each student understand that the responsibilities associated with high standards of excellence include ensuring that all class work undertaken in this program is performed in an environment that promotes serious scholarship and moral rectitude. Though only summarized here, this class herein delineates a zero-tolerance policy for acts of academic dishonesty. All acts of suspected academic dishonesty will be thoroughly investigated in a manner that is fair, timely, and efficient and done so in a manner that protects the rights of both faculty member and student, in meeting and following Boston University standards and protocols. Any individual who is found to have committed an act of academic dishonesty may receive a penalty, up to and including expulsion from Boston University.

The official Boston University code of conduct as well as its statement on academic dishonesty is available in its entirety online at <http://www.bu.edu/academics/resources/academic-conduct-code/>.

Students are expected to be fully aware not only of all expectations but also consequences for violations. Additional questions about appropriate academic conduct should be brought by students to their course instructor, primary advisor or the Program Director *before*, not after, work is submitted.

### *Plagiarism*

Self-plagiarizing is not allowed in any circumstance, which means that students are not permitted to submit their own work that was already submitted in any other coursework. All assignments must refer carefully to the sources used. Copying the ideas and results of other authors (either word for word, or as a paraphrase) without explicit reference to the source is considered to be plagiarism.

It is your responsibility to familiarise yourself thoroughly with the faculty's policy on unfair practices, fraud and plagiarism.

Seriously, this is bad. Really bad. If you are unfamiliar with recent high-profile cases of academic dishonesty, please read about [Bauman](#) and [LaCour](#).

### *Feedback*

Feedback will be given regularly in class, on weekly assignments, and by appointment.