

*Boston University College of Communication*

*Division of Emerging Media Studies*

Syllabus

**EM747: #Trending Insights: Social Data Analysis   
and Visualization**

Instructor:

Jacob Groshek, Ph.D.

1. Practical Information

|  |  |
| --- | --- |
| **Course Name** | #Trending Insights: Social Data Analysis and Visualization |
| **Course Code** | EM747 |
| **Credits**  **Term** | 4  Spring 2015 |
|  |  |
|  |  |
| **Instructor** | Dr. Jacob Groshek |
|  | 704 Commonwealth Ave.  302D |
|  | [jgroshek@bu.edu](mailto:jgroshek@bu.edu) |
|  | 617-353-6421 |
|  |  |
|  |  |
| **Office Hours** | 10:30am until 12:30pm on Tuesdays and Thursdays  2:00pm until 3:00pm on Tuesdays and Thursdays  Also by appointment |
|  |  |
| **Location** | 704 Commonwealth Ave.,  Room B04 |
|  |  |
| **Timetable** | Section A1 – 12:30pm until 2:00pm Tuesdays and Thursdays |
|  |  |
|  |  |
| **Course Website** | dropbox.com/EM747  gorilladragon.com |
|  |  |
|  |  |
|  |  |
| **Compulsory Literature** | Steele, J. & Iliinsky, N. (2010). *Beautiful Visualization.* Cambridge, MA: O’Reilly Media. ISBN: 978-1-449-37987-2. (e-book / kindle edition recommended) |
| **Optional Literature** | *Additional readings are available on dropbox, the Internet, or via the BU library.* |
|  |  |

1. Course Overview

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Wk.** | **Topic** | **Literature** | **Software** | **Outcome** |
| 1  (1/20 -22) | Introduction to Course and Introduction to Software | DataScience, Ch1-2 | Install Gephi and Filezilla | Demonstrate comfort with data, successfully install and use software |
| 2  (1/27-29) | Collecting and Cleaning (w/ Stu Shulman, Discovertext founder) | KKV, Ch1; Lu & Shulman, 2008; McQueen et al., 1998; Morse, 1995 | Discovertext | Search and collect original social data into archives |
| 3  (2/3-5) | SNOW DAYS | -- | -- | -- |
| 4  (2/10-12) | Analytic and Theoretical Background | BV, Ch 1; Rogers, 2013; Harris, 2013; Zizi, Intro | Web-based examples | Demonstrate command of conceptual understanding of data visualization; map own network |
| 5  (2/17-19) | Client Workshop  **(No Class on Tuesday, 2/17)** |  | TCAT; GorillaDragon | Successfully identify real-world client |
| 6  (2/24-26) | Visualizing other Social Networks | BV, Ch7; BV, Ch10; Knight Lab, 2014; Zizi, Ch5; Gieseking, 2013 | Nevizz; Gephi | Demonstrate competence with graphical tools and datasets; map networks and find communities with Gephi |
| 7  (3/3-5) | Data Mining Basics | BV, Ch11; O’Connor et al., 2010; Knibbs, 2013; Zizi, Ch4 | Discovertext; TCAT | Mine social data for certain users; export data for sorting and analysis |
| 8 (3/10-12) | Spring Recess, Classes Suspended | Eat | Drink | Be Merry |
| 9  (3/17-19) | Static Visualizations of User Networks | Groshek & al-Rawi (2013); BV, Ch9; Reider, 2012 | TCAT; Gephi | Create image-based user network graphs; weight output |
| 10  (3/24-26) | Dynamic Visualizations of User Networks | BV, Ch19; Bruns, 2012; Freelon, 2013 | TCAT; Gephi | Create web-based interactive user network graphs; weight output |
| 11  (3/31-4/2) | Dynamic Visualizations of Concept Networks | BV, Ch19; Rosling, 2007/9; Goodman, 2009 | TCAT; Gephi | Create web-based interactive concept graphs; weight output |
| 12  (4/7 – 9) | Measuring Sentiment | Thewall et al., 2011, 2012; Dang-Xuam et al., 2013; Vargo et al, 2013 | TCAT; Excel; Sentistrength | Extract and model sentiment amongst unique user groups |
| 13 (4/14-16) | Data Mining and Time | BV, Ch2; Gerlitz & Rieder, 2013; Papacharissi & Oliveira, 2012 | TCAT; Excel | Code and validate date; Visualize trends in content production over time |
| 14  (4/21-23) | Visualizing Geospatial Relationships | BV, Ch6; NYTimes, 2013; FlowingData, 2013: Tsou & Leitner, 2013 | TCAT; Gephi | Create (cautious) geolocative representations of data |
| 15  (4/28 – 30) | Final Project Workshop | Student-supplied project specific readings | Any; All | Develop models to explore, see, and answer questions |

1. Course Introduction

This course familiarizes students with social-scientific methods for large-scale data analysis and visualization, including the application of relevant user and concept networks, time and spatial models, and sentiment analysis. In addition, the use of germane software in emerging and digital media research is developed.

More importantly, however, this course has a dual structure where students learn to not only carry out advanced analyses of large datasets, they also engage with how to visually represent and effectively communicate those results to a lay audience. As such, students leave equipped with a wide-ranging skillset to scrape data, mine data, and present data in fields of specific areas of inquiry.

1. General Course Objectives

*Students have the knowledge and understanding of:*

1. How to use build large datasets (100,000+ units) using popular social media (Facebook, Twitter, and relevant other) platforms;
2. how to engage with various software systems to collect and analyze data along specific dimensions of topic and influence;
3. how to use open source software to visualize networks and concept patterns in large social datasets;
4. where there are benefits and limitations of these approaches, and how those can be minimized or complemented in telling interesting, visual stories with data.

*Students have the ability to:*

1. Harvest social media data into large datasets for analysis;
2. code, validate, and adjudicate human decisions about social media content;
3. use human coding to train machine learning to code content at scale;
4. map networks of users to visualize human network structures;
5. map networks of concepts to visual semantic network structures;
6. develop time-series models that explain and predict outcomes from social data;
7. measure sentiment in big data with humans and/or machines;
8. weight content by measures of influence to model information flows;
9. develop appropriate analytic techniques to mine data for significant patterns;
10. apply various software interfaces for both analysis and visualization;
11. carry out original research that presents visual data relationships with efficiency, informativeness, novelty, and (perhaps most of all) beauty.
12. Organization & Working Method

General

It is necessary to attend class sessions because without them there is no course. Therefore it is 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.

Web-based Readings

The majority of readings will be made available in online format, including the textbook that is available as an e-book through Amazon or the publisher. It is worth noting that this class regularly relies on exercises based online. Additional readings from journals or elsewhere will be accessible through URLs or through dropbox.

Rules about absence

Students who have a serious reason for missing a meeting should notify the instructor in advance by email or telephone if they so choose. No extra credit will be (probably be) offered, and there are no additional make-up assignments.

1. Assessment and Grading

Students will complete 11 weekly assignments in total, all of which are to be completed individually, except in rare cases. Students are, however, welcomed to collaborate in solving problems, but each must work independently with unique data for each assignment.

The deadline for all assignments 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 assignments**

The assignments are all a combination of empirical and practical research considerations where students must seek to answer substantive research questions. Each assignment asks students to apply the discussed readings and carry out some form of analysis and/or visualization toward the end of advancing knowledge, not just visualizing for the sake of it.

We will be doing work for real-world clients in this class, and we will have a working portfolio at the website of the online consultancy gorilladragon.com

Each assignment will be laid out in full during the weekly class session.

**Criteria and grading**

*Criteria used to evaluate 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 as well as contributing fully and equally to the assignments.  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  90-92.99 A-  87-89.99 B+  83-86.99 B  80-82.99 B-  77-79.99 C+ | 73-76.99 C  70-72.99 C-  67-69.99 D+  63-66.99 D  60-62.99 D-  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 | 40 | Team (optional) | 20.0% |
| Final Presentation | 30 | Team (optional) | 15.0% |
| Assignments (x 11) | 10 (x11 = 110) | Individual | 5.0% (x11 = 55%) |
| Participation | 20 | Individual | 10.0% |
| Total | 200 |  | 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*

The assignments and final project produced in the class are team products. It is not allowed to use work from other teams. It is permitted, though, to discuss each other's work. 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.

The submission of electronic versions of the assignments in Blackboard’s SafeAssign is necessary to facilitate (automatic) checks on plagiarism. It is your responsibility to familiarise yourself thoroughly with the faculty’s policy on unfair practices, fraud and plagiarism.

*Feedback*

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