MPJO-722-01: DATA REPORTING

GEORGETOWN UNIVERSITY: MPS-JOURNALISM

Thursdays, 5:20 p.m. to 7:50 p.m. | Spring 2015

Instructor: Derek Willis
Downtown campus, room C226 (Mac lab)
- Office hours are by appointment.

COURSE OVERVIEW

The world that we live in and report on is increasingly influenced by data. This course will help you use data to find and develop stories that can't be found in other ways.

Data can be anything: spreadsheets full of numbers, the text of speeches or the measured observations of daily life. Reporters need to be able to treat data as another source to be researched, interviewed and analyzed, using the right tool for the job. Beginning with spreadsheets and continuing to databases, basic mapping and rudimentary programming, this class will make working with data a part of your skill set.

This is a skills-based course, so students will need to be comfortable with learning to use computer software beyond word processors. If you are wondering if you can do this stuff, you can. You may prefer interviewing people to data, but avoiding data - particularly for Washington journalists - is no longer an option.

COURSE OBJECTIVES

By the end of the semester, students will have:
- learned how to interview data as they would a human source
- used different software and Internet utilities to help find stories within data
- surveyed and discussed data journalism projects
- completed a final project involving the acquisition and analysis of data

REQUIRED READING

There is no required text for this course. Instead, the instructor will assign weekly readings chosen from professional data work, government documents, blog posts and other sources. Students will be expected to discuss these readings in class.

ATTENDANCE

Journalism is not a passive activity and requires focus, inquisition and involvement. We will be discussing professional work, writings and data issues every week, and I expect your comments, questions and other contributions to our class. None of this can happen if you don't show up. If you must miss a class, please try to let the instructor know in advance. You will be responsible for any work you miss.

As outlined by the university, missing more than two classes will result in a final grade reduction of one level (for example, an A will be converted to an A-). **Absences for classes beyond the initial two will result in further reduction of the final grade.** If you are absent for more than four classes, you will be in danger of failing this course.

CLASSROOM ETIQUETTE AND OTHER GUIDELINES

Respect: You should treat your classmates with respect. I'll expect this both in your verbal communication but also in your non-verbal communication. This means: Pay attention and be empathetic.

Participation: Classes are long. We'll take a break each class, but I expect you to be attentive while we're in session. This means: If we're having discussion, you're participating. If we're working on our projects, you're engaged. Also: Park your phone and messaging apps unless instructed to use them for class purposes.

Late work: Late work will be penalized on a sliding scale (the later the assignment, the larger the penalty).

Instructional continuity: In the event of a weather emergency (or any other widespread emergency) that would close the Georgetown Downtown building, we will plan to meet virtually through online videoconferencing tools or at another location if possible. More information will be provided on how this will work later in the semester.

ASSIGNMENTS

Homework: Students will be required to complete exercises involving the use of spreadsheets, databases and other tools both during class and outside of class. It is your responsibility to find computer time for the outside assignments. A good rule: Don't come to class empty-handed – at the very least, show me that you attempted the assignment.
Deadline: Beginning of class (or emailed before class)

Critiques: Each student will complete two written critiques of professional work that makes extensive use of government data. You may not be able to replicate the reporters' work, but try to put yourself in their shoes and judge the decisions they made. Think of these as mini-book reviews, and I expect between 500 and 750 words for each.
Deadline: Jan. 29 & Feb. 19

Midterm data analysis: Each student will be responsible for obtaining, analyzing and visualizing a federal government data set assigned to him/her, using methods we'll cover in class.
Date: March 5

Story Memo: Students will obtain and analyze federal government data and write a story memo about their findings, rather than writing an actual story. The emphasis is on the process

more than the final product, although the memo should reflect the depth of your work. Some in-class project time will be provided, but prepare to spend significant time outside class working on this assignment. Each student will be required to hand in a well-written memo describing in detail the work done on the project and the handling of the underlying data. The memo should also address any weaknesses in the data or unexpected events that hampered or improved the process. The memo takes the form of a story pitch to an editor, meaning they should be long enough to both make the case (or lack thereof) and demonstrate the analysis the student has done. Aim for at least 1,000 words, and supplemental materials are welcomed.
Proposal Deadline: April 9
Final Deadline: May 7

Data Assessment: Each student will, as part of the story memo project, undertake a study of the data offerings of the agency that produced the data set used for the story memo. This study will result in a short paper of between 500 and 750 words assessing the scope, quality and accessibility of the agency's data, how it has been used by journalists and how it could be used. The paper should also identify opportunities for agencies to release or organize data of public interest. Consider it like a critique, but of an agency rather than a story.
Deadline: May 7

GRADING
Your course grade will be based on the following:

Weekly homework:	25%
Critique 1:	5%
Critique 2:	5%
Midterm data analysis:	20%
Story memo:	25%
Data assessment:	5%
Attendance/participation:	15%
Total:	**100%**

Graduate course grades include A, A-, B+, B, B-, C and F. **There are no grades of C+, C- or D.**

A	100-93		B-	82.99-80
A-	92.99-90		C	79.99-70
B+	89.99-88		F	69.99-0
B	87.99-83			

The instructor will provide a warning by mid-semester to any student who appears to be on track for a poor final grade.
UNIVERSITY RESOURCES

Georgetown offers a variety of support systems for students that can be accessed on main campus or at the downtown location:

- MPS Writing Resource Program
  202-687-4246
  http://writingcenter.georgetown.edu/

- Academic Resource Center
  202-687-8354 | arc@georgetown.edu
  http://ldss.georgetown.edu/

- Counseling and Psychiatric Services
  202-687-6985
  http://caps.georgetown.edu/

STUDENTS WITH DISABILITIES POLICY

Students with documented disabilities have the right to specific accommodations that do not fundamentally alter the nature of the course. Students with disabilities should contact the Academic Resource Center (Leavey Center, Suite 335; 202-687-8354; arc@georgetown.edu; http://ldss.georgetown.edu/index.cfm) before the start of classes to allow time to review the documentation and make recommendations for appropriate accommodations. If accommodations are recommended, you will be given a letter from ARC to share with your professors. You are personally responsible for completing this process officially and in a timely manner. Neither accommodations nor exceptions to policies can be permitted to students who have not completed this process in advance.

GEORGETOWN HONOR SYSTEM

All students are expected to maintain the highest standards of academic and personal integrity in pursuit of their education at Georgetown. Academic dishonesty in any form is a serious offense, and students found in violation are subject to academic penalties that include, but are not limited to, failure of the course, termination from the program, and revocation of degrees already conferred. All students are held to the Honor Code. The Honor Code pledge follows:

In the pursuit of the high ideals and rigorous standards of academic life, I commit myself to respect and uphold the Georgetown University Honor System: To be honest in any academic endeavor, and To conduct myself honorably, as a responsible member of the Georgetown community, as we live and work together.

PLAGIARISM

Stealing someone else's work is a terminal offense in journalism, and it will wreck your career in academia, too. Students are expected to work with integrity and honesty in all their assignments. The Georgetown University Honor System defines plagiarism as "the act of

passing off as one's own the ideas or writings of another." More guidance is available through the Gervase Programs at http://gervaseprograms.georgetown.edu/honor/system/53377.html. If you have any doubts about plagiarism, paraphrasing and the need to credit, check out http://www.plagiarism.org.

SYLLABUS MODIFICATION

The syllabus may change to accommodate discussion of emerging topics. Also, the schedules of guest speakers may require some shifting of our agenda. The instructor will make every effort to provide as much advance notice as possible for any alterations.

CLASS SCHEDULE

WEEK 1 (THURSDAY, JANUARY 8)
- Introductions
- What is data?
- Spreadsheet skills

WEEK 2 (THURSDAY, JANUARY 15)
- Unstructured data
- Working with text data
- Basic visualizations

WEEK 3 (THURSDAY, JANUARY 22)
- Excel formulas
- Basic charting
- Election results data

Assignment: Critique No. 1
Deadline: Beginning of class, Jan. 29

WEEK 4 (THURSDAY, JANUARY 29)
- Introduction to SQL and SQLite
- Campaign finance data

WEEK 5 (THURSDAY, FEBRUARY 5)
- Summarizing data using SQL
- Fusion Tables
- More campaign finance data

WEEK 6 (THURSDAY, FEBRUARY 12)
- SQL joins
- Congressional voting data

Assignment: Critique No. 2
Deadline: Beginning of class, Feb. 19

WEEK 7 (THURSDAY, FEBRUARY 19)
- Fuzzy SQL
- The other parts of SQL
- More congressional data

WEEK 8 (THURSDAY, FEBRUARY 26)
- Guest speaker: TBA
- NICAR conference

Assignment: Midterm data analysis
Deadline: Beginning of class, March 5

WEEK 9 (THURSDAY, MARCH 5)
- Cleaning dirty data
- Mapping

THURSDAY, MARCH 12 – NO CLASS (SPRING BREAK)

WEEK 10 (THURSDAY, MARCH 19)
- Working with PDFs
- Military data

Assignment: Story memo proposal
Deadline: Beginning of class, April 9

WEEK 11 (THURSDAY, MARCH 26)
- Mapping & Census data
- Writing and editing with data

THURSDAY, APRIL 2 – NO CLASS FOR EASTER BREAK

WEEK 12 (THURSDAY, APRIL 9)
- Structured Web data
- Pattern matching
- IRS data

WEEK 13 (THURSDAY, APRIL 16)
- Basic programming
- Transportation data
- In-class project work

WEEK 14 (THURSDAY, APRIL 23)
- Real-time data
- Using APIs
- In-class project work

THURSDAY, APRIL 30 – NO CLASS (STUDY DAYS)

WEEK 15 (THURSDAY, MAY 7)

- Project presentations
- Wrap-up