Syllabus: JOMC570 Data Driven Journalism - Spring 2020

Ryan Thornburg

January 16, 2020

MW – 9:30-10:45 a.m. (Section 1)

MW - 12:30-1:45 p.m. (Section 2)

Instructor: Ryan Thornburg

Office Hours: 8 a.m. - 4 p.m. Tues/Thurs

Office Location: CA 215

Office Phone: 919-537-3681

Course Site: https://rtburg.github.io/MEJO570-Spring2020/

PRE-REQUISITES

Experience reporting news in a professional or classroom setting, including clear news judgment and ability to communicate with clarity and brevity.

COURSE DESCRIPTION

In this skills-based course students learn how to acquire, clean, analyze and present data in a journalistic setting.

Your decision to take this course indicates that you are interested in learning the skills and concepts of data-driven reporting. The class starts from the assumption that you've never or rarely used even a basic spreadsheet to aid either your reporting or storytelling. That's where the semester will begin. Along the road to data literacy we will also go over some basic statistics and basic data visualization concepts.

OBJECTIVES

Students who successfully complete this course will be able to acquire, organize, analyze and present data to a general news audience. Students will learn how to use the tools of data-driven journalism as a means to developing and testing hypotheses that lead to transparent and reproducible data-driven stories.

Through the study of quantitative reasoning and methods, students in this course will acquire and reinforce the ability to use analytic and quantitative ideas as they are applied in the context of professional journalistic reporting. In today's world of fast-paced advances in the use of data, the importance of such skills cannot be overstated.

This course focuses especially on the ways that quantitative reasoning can be applied to the kinds of data that professional news reporters use in a variety of media presentation formats.

Students in this course will learn how to collect and interpret quantitative data, apply mathematical analysis in a news reporting context and use numerical reasoning to organize and present a story to general news audiences.

ATTENDANCE POLICY

As part of their grade, students will submit after each class session one question they wished they had asked about the material covered in class. These submissions will be used to track attendance and will be part of the "daily work" grade, along with homework exercises and reading summaries described later. Students who do not attend a class session will not be allowed to submit a question about the session they missed.

For each class, I will assign one student to take lecture notes that will be shared with other members of the class via Sakai. For the class sessions where we are working in R (starting with Class 5) students will take the lecture notes in an R notebook file. On the day you take notes you will not be responsible for a question.

LATE ASSIGNMENTS

In general, I need you to turn in your assignments on time so I can give you feedback in time for it to be useful to you.

For homework assignments, even 1 second after deadline is -1 point. After 24 hours, the deduction becomes 2 points. The grade for a homework assignment that is 48-72 hours late will be reduced by 3 points. I will not accept homework assignments more than three days late unless we have made arrangements before deadline.

For major assignments, students can miss deadlines on major assignments by a cumulative 48 hours without penalty. That means one assignment can be 48 hours late, or 4 assignments can each be 12 hours late, etc. Once the 48-hour grace period is expended, the student will immediately begin losing 10% of the possible points for each fraction of a day the assignment is late. Grades on late assignments will immediately drop a full letter grade each calendar day it is late (eg. An A assignment due at noon on Thursday becomes a B at 12:01 p.m., a C at noon on Friday, a D at noon on Saturday, and an automatic F after noon on Sunday.)

COURSE MATERIALS

Your laptop is the tool of your craft, and its care and maintenance reflect upon you and your readiness to work in the professional settings for which this course prepares you. Just like a guitar player wouldn't show up to rehearsal with a broken string, you also need to ensure your computer is in good working order each and every class. Laptops – like guitar strings – will break. Having a backup plan demonstrates maturity and professional preparedness.

By the second day of class you will need to have installed on your laptop several free software applications we will be using throughout the semester.

All of the reading material you need for class is either free online, free via the generosity of UNC Libraries, or free via electronic reserves at UNC Libraries.

I strongly encourage you to purchase a $25 IRE student membership (https://www.ire.org/membersonly/join/register) . This will give you access to dozens of tip sheets and resources that will be useful to you in class and as part of your broader professional development.

DETERMINATION OF GRADE

Throughout the semester, students will earn points for completing various assignments. Most of the assignments will be given letter grades that correspond to the numerical values found on the UNC Registrar's site. When determining final grades for the course, I will trim the values to the tenths place. (For example, a 3.7 will be an A-, but a 3.70001 will be an A.)

Assignment	Percentage of your grade	Due date
Attendance questions	10%	Each class

Assignment	Percentage of your grade	Due date
Homework and quizzes	40%	As assigned
Data negotiation	10%	April 15
Final	40%	See below

Final project

For the final project you will be learning how to find and report stories in public political campaign finance data by examining reports for candidates who are running for state office in North Carolina this year.

We will begin work on the final project in Week 7, and it will have portions due throughout the rest of the semester. The final project is due on the final exam date for your section.

More about the final project

Extra Credit

In addition to the required assignments, there will be two opportunities to earn extra credit.

If you produce **a published news story that relies on data you receive as part of your records request**, you may earn extra credit of up to 10% of your final grade. Thie story must be published by the last day of class.

You may also complete the *"How I got the story"* extra credit assignment for up to 10% of your final grade. This must be done by Spring Break.

Additionally, every typo, broken link or other error you find in the syllabus before the start of the third week of class will earn you one extra credit point on your participation grade. Send all errors to me via email.

Passing Grades

If you are a MEJO major using this to satisfy a MEJO course requirement, you must receive at least a C- in the course in order to have the credits count toward your major requirements.

If you are a MEJO minor, you must receive at least a C in the course in order to have the credits count toward your minor.

Please see "Undergraduate Grade Definitions" and "Graduate Grade Definitions" at https://registrar.unc.edu/academic-services/grades/explanation-of-grading-system/ for a better understanding of what letter grades mean at UNC-Chapel Hill.

STUDENT HONOR

The University of North Carolina at Chapel Hill has had a student-led honor system for over 100 years. Academic integrity is at the heart of Carolina and we all are responsible for upholding the ideals of honor and integrity. The student-led Honor System is responsible for adjudicating any suspected violations of the Honor Code and all suspected instances of academic dishonesty will be reported to the honor system.

All academic work in this course, including homework, quizzes, and exams, is to be your own work, unless otherwise specifically provided. It is your responsibility if you have any doubt to confirm whether or not collaboration is permitted. If the work is truly your own, you will be able to explain and demonstrate to my satisfaction how you did it.

Do not represent someone else's words, thoughts, or ideas as your own without attribution in connection with submission of academic work, whether graded or otherwise.

If you have any questions about your responsibility or your instructor's responsibility as a faculty member under the Honor Code, please see the course instructor or Senior Associate Dean C. A. Tuggle, or you may speak with a representative of the Student Attorney Office or the Office of the Dean of Students.

Further information about the student Honor Code is available at http://honor.unc.edu

SEEKING HELP

If you need individual assistance, it's your responsibility to meet with the instructor. If you are serious about wanting to improve your performance in the course, the time to seek help is as soon as you are aware of the problem – whether the problem is difficulty with course material, a disability, or an illness.

Before you meet with me to help you with code, you need the following so I can best help you:

- A description of the problem (using Stack Overflow's guide)
- A description of what you've done to solve the problem yourself
  - What did you do to try to isolate the error?
  - What search phrase did you use and why did that not work?
  - Did you have a peer help you try to diagnose?
- A copy of the code and data you are using.
- An appointment to meet in office hours

I will *always* help you. But I need to see that you're working just as hard as I am to find the solution. Also, I expect you to make mistakes. We all make mistakes. You will see me make these kinds of mistakes in class. But there are three kinds of mistakes that if I find during a troubleshooting session with you will cost you a point on your next homework assignment. They are:

- Spelling errors
- Capitalization errors
- Missing (or too many) enclosure marks (like quote marks, parentheses, etc.)
- Spacing or line break errors

Why so mean to take off a point for these kinds of errors? Because they are errors of care and attention, not understanding. And because they are easy to eliminate as possible sources of error.

DIVERSITY

UNC is committed to providing an inclusive and welcoming environment for all members of our community and does not discriminate in offering access to its educational programs and activities on the basis of age, gender, race, color, national origin, religion, creed, disability, veteran's status, sexual orientation, gender identity, or gender expression. The full University policy is available from the office of Equal Opportunity and Compliance.

ACCOMODATIONS

If you require special accommodations to attend or participate in this course, please let the instructor know as soon as possible. If you need information about disabilities visit the Department of Disability Services website at http://disabilityservices.unc.edu/

ACCREDITATION

The School's accrediting body outlines a number of values you should be aware of and competencies you should be able to demonstrate by the time you graduate from our program. Learn more about them here: http://www2.ku.edu/~acejmc/PROGRAM/PRINCIPLES.SHTML#vals&comps

No single course could possibly give you all of these values and competencies; but collectively, our classes are designed to build your abilities in each of these areas. In this class, we will address a number of the values and competencies, with special emphasis on the last six bullet dots under "Professional values and competencies" in the link above.

COURSE CALENDAR

Week 1

Class 1: Jan. 8

In class:

- Getting into a data state of mind
  - Miami police officer arrested for speeding
- Syllabus
- Introductions

Homework:

- Read: Data Literacy, Chapter 1
- Read: The New Precision Journalism, Chapter 1
- Read: What is data?
- Read: Why journalists should use data
- Read: Why is data journalism important?
- Read: Data journalism in perspective
- Read: How the Sun Sentinel reported its Pulitzer Prize winning coverage of off-duty cops
- Do: Reading summaries 1
- Do: Complete computer set-up assignment

Week 2

Class 2: Jan. 13

In Class:

- Getting into a data state of mind with spreadsheets
- Basic Spreadsheets: rows, columns, cells, importing, sorting, filtering

Homework:

- Do: Data state of mind assignment
- Read: Data Literacy, Chapter 2
- Read: Data Literacy, Chapter 3
- Do: Reading summaries 2

Class 3: Jan. 15

In Class:

- Public records lecture
- Introducing public records negotiations diary assignment
- Quick assignment of candidates and class note dates

Homework:

- Read: Data Literacy, Chapter 4
- Read: Art of Access, Chapter 7 (available via UNC Libraries digital reserves
- Do: Reading summaries 3

Week 3

Jan. 20: MLK Day. No Class

Class 4: Jan. 22

In class:

- Intro to data analysis with spreadsheets: What's normal and not normal?
- Excel functions: mean, median, max, min, rank

Homework

- Do: Excel Assignment 1

Week 4

Class 5: Jan. 27

In class:

- How data can lie to you
- Introduction to R and RStudio
  - SettingUpForRClasses.R
- Introduction to markdown and R notebooks
- Installing R packages
- Introduction to variables, operators, lists, functions

Homework:

- Read: Evolve from Excel to R
- Watch: R in the data journalism workflow at FiveThirtyEight

- Skim: R for Data Science, "Workflow: basics"
  - Note for here and elsewhere: The printed version of "R for Data Science" has different chapter numbering, so the names of the chapters are what you should pay attention to.)
- Skim: R for Data Science, "Workflow: projects"
- Skim: R for Data Science, "R Markdown"
- Skim: R for Data Science, R Markdown Workflow
- Skim: R for Journalists, "How to use R"
- Do: R Assignment 1. Workspace set-up and first R notebook

Class 6: Jan. 29

In Class

- Loading MLB Salaries Data into R
- Installing and loading packages
- First steps with data analysis in R

Homework

- Do: R Assignment 2. First steps with data analysis in R using MLB Payrolls
- Read: R for Journalists, "Transforming and Analyzing Data"
- Read: R for Data Science, "Data Transformation with Dplyr (Sections on filter, arrange and select)"

Week 5

Class 7: Feb. 3

In Class:

- Introduction to Tidyverse and Dplyr packages
- comparison operators, logical operators, missing values, filter(), arrange(), select(), rename()
- Filtering data: Imagine an interview where you can only ask "yes" or "no" questions.
- Computers are hardworking and obient, but they just don't get you like I do.
- How filtering data is like writing a good records request.
- Work together on replicating MLB Salaries assignment in R

Homework

- Do: HW6. Using filter(), arrange() and select() in R
- Read: R for Data Science, "Data Transformation with Dplyr (Section on mutate)"

Class 8: Feb. 5

In Class:

- Creating new columns with the mutate() function in R
- Work on HW7. Mutate together in class

Homework

- Complete HW7. Mutate
- Read: What is a data state of mind, and how can you develop it?
- Watch: Data state of mind
- Watch: Numbers in the Newsroom
- Do: Reading/watching summaries 4

Week 6

Class 9: Feb. 10

In Class:

- Back to Excel: Summarizing data with pivot tables

Homework

- Do: HW8: Excel Pivot Tables
- Read: R for Data Science, "Data Transformation with Dplyr (Section on group_by and summarize)"
- Read: R for Journalists,"Tidying and joining data" (just the spread and gather sections)
- Read: R for Data Science, "Tidy data" (just 12.1-12.3)

Class. 10: Feb. 12

In Class:

- Doing pivot tables in R with group_by(), summarize() and spread()
- Work on start of HW9 in class

Homework:

- Do: Analyzing UNC Salaries in R
- Read: R for Data Science, "Data import"
- Read: R for Journalists, Chapter 2

Week 7

Class 11: Feb. 17

In Class:

- Review group_by(), summarize() and spread()
- Importing data into R
- Dealing with different filetypes
- Changing data types
- Renaming columns
- Dealing with missing variables

Homework:

- Do: Load your candidate's data
- Read: R for Data Science, "Data visualization"

Class 12: Feb. 19

In Class:

- Introduction to exploring data through visualization

Homework

- Read: Numbers in the Newsroom, Chapter 2
- Do: Reading summaries 5
- Do: R for Data Science, 3.2.4 Exercises 1,2,4,5
- Do: R for Data Science, 3.3.1 Exercises 1,2,3,4
- Do: R for Data Science, 3.5.1 Exercises 1-6
- Do: R for Data Science, 3.6.1 Exercises 1-6
- Do: R for Data Science, 3.9.1 Exercise 2

Week 8

Class 13: Feb. 24

In Class:

- Continuous vs. categorical variables in visualization
- Visually exploring campaign finance reports
- Discuss Final Assignment: Analysis, Draft, Final
- Choose your candidates for final project
- Doing a clips search
- Finding experts

Homework

- Explore: data visualization options with "from Data to Viz"
- Read: Statistics for the Newsroom
- Read: R for Data Science, "Exploratory Data Analysis" (through "Covarition Two categorical variables")

Class 14: Feb. 26

In Class:

- Working with strings in campaign finance data

Homework:

- Read: R for Journalists, “Handling Strings”
- Read: R for Data Science, “Strings”
- Do: “Handlings strings” exercises in R for Journalists" Optional:
- Do: R for Data Science, 14.3.3.1 Exercises 1, 5
- Do: R for Data Science, 14.3.4.1 Exercises 1-4
- Do: R for Data Science, 14.3.5.1 Exercises 1-2
- Do: R for Data Science, 14.4.4.1 Exercises 1-2
- Do: R for Data Science, 14.4.3.1 Exercises 1-2
- Do: R for Data Science, 14.4.5.1 Exercises 1-3

Week 9

Class 15: March 2

In Class:

- Working with dates in campaign finance data

Homework:

- Read: R for Journalists, “Dealing with Dates”
- Read: R for Data Science, “Dates and times”
- Do: “Dealing with Dates” exercises in R for Journalists"

Class 16: March 4

In Class:

- More useful R tricks:
  - ifelse()
  - case_when()
  - separate()
  - unite()
  - cast()
  - gather()

Homework:

- Read: R for Data Science, Chapter 9
- Read: R for Journalists, Tidying and joining data

March 9 - 13: Spring Break

Week 10

Class 17: March 16

In Class

- Due: "How I got the story?" extra credit assignment
- Introduction to Joins.

Homework:

- (Re-)Read: R for Journalists, Tidying and joining data
- Read: R for Data Science, "Relational Data with dplyr"

Class. 18: March 18

In class:

- More joins.

Homework:

- Do: Join campaign finance to census data

Week 11

Class 19: March 23

In Class:

- Open Refine Day 1

Class 20: March 25

In Class:

- Open Refine Day 2

Homework:

- Read: Numbers in the Newsroom, Chapter 1
- Read: Numbers in the Newsroom, Chapter 6
- Do: Reading summaries 6

Week 12

Class 21: March 30

In Class:

- DUE: Data Analysis for Final Story
- Dealing with data in PDFs and HTML
- CometDocs & Tabula
- ImportHTML

Class 22: April 1

In Class:

- ImportXML

Homework:

- Do: HW 14: Scrape IRE jobs board using ImportXML (instructions on Sakai)

Week 13

Class 23: April 6

In Class:

- Intro to Rvest

Class 24: April 8

In Class:

- Rvest 2

Week 14

Class 25: April 13

Homework:

- Read: Numbers in the Newsroom, Chapter 3
  - Do: Reading summaries 7
- Read: R for Journalists, Chapter 4
- Read: R for Data Science, "Graphics for communication with ggplot"

Extra Credit Reading:

- https://ggplot2-book.org/
- https://rstudio.com/wp-content/uploads/2015/03/ggplot2-cheatsheet.pdf
- http://www.cookbook-r.com/Graphs/

Class 26: April 15

In Class:

- Final data negotiation diaries due
- Mapping campaign finance

Homework

- Read: R for Data Journalists, Chapter 5

Week 15

Class 27: April 20

In Class:

- Panel discussion with data reporters

Class 29: April 22

In Class:

- Final Story Drafts due at start of class.
- Posting your work to github

Final Dates

Section 2: Tuesday, April 28, 12 p.m.

- Final Story Edits and meet with professor
- Present data negotiations war stories
- Final Story due at 3 p.m.

Section 1: Monday, May 4, 8 a.m.

- Final Story Edits and meet with professor
- Present data negotiations war stories
- Final Story due at 11 a.m.