Course Title:	**Computational Research Methods (using R)**
Course Number:	MEJO 890.2
Instructor:	Prof. Deen Freelon, Ph.D.
Time:	W 2 pm – 445 pm
Room:	Curtis 303
Office hours:	Remote and by appointment only.
	Schedule a Zoom appointment with me
My email:	freelon@email.unc.edu
Course website:	https://sakai.unc.edu/portal/site/mejo890.001.sp23

Course introduction

Millions of people use social media every day. Making sense of all that content is one of the great challenges of the early 21st century. The skills needed to conduct such research are prized not only by social science researchers like me, but also by companies who want to know how customers are responding to their products and nonprofits interested in measuring the efficacy of their cause campaigns. Students in this course will learn computer programming skills and apply them to an individual research project that analyzes, visualizes, and draws meaningful conclusions about social media data. By the end of the course they will be prepared to conduct basic computational social science research as well as continue their computational education in other classes or independently.

By the end of this course, you should be able to:

- Preprocess and standardize Twitter and/or Crowdtangle data using the R programming language
- Conduct basic descriptive social media analysis (see below)
- Visualize research results
- Explain clearly how research findings relate to practical and/or theoretical concerns
- Understand key theoretical and methodological issues in computational communication research

Summary of course requirements

- Attend and participate in all class meetings – 10%
- Coding homework – 40%
- Final research project and presentation - 50%

I also offer an undergraduate version of this course that teaches some of the same skills (MEJO 571). The table below summarizes the main similarities and differences between the two courses:

	MEJO 890 (grad course)	MEJO 571 (undergrad course)
No prerequisites	x	x
Python-based		x
R-based	x	
Final project requires production of original code	x	x
Final project requires extended, theoretically grounded writeup	x	
Required readings explore theoretical, methodological, and practical issues in computational research	x	
Accelerated instructional pace	x	
Class sessions	Mix of coding labs and discussion seminars	All coding labs
Meeting schedule	1x weekly	2x weekly
Enrollment restrictions	MA and PhD students only	Undergrads and MA students

Detailed course plan

This course incorporates two distinct pedagogical tracks that run concurrently. The first track's goal is to teach you *how* to use computer programming as a research method. To that end, we will spend most of each class session writing, reviewing, and debugging R code. You will complete a code-based homework assignment most weeks to help you hone your skills. Most of these will be fairly short, but several will be longer.

The second track is more like a traditional seminar, although the reading load is somewhat less intense. While the coding track focuses on *how* to use code for research, the readings address *why* code has value for conducting social science research and *what* are the best ways to use it to generate high-quality research. Part of most class sessions will be devoted to discussing the readings.

The exact amount of coding material we cover in the *how* track will depend on how quickly you all can master it. Different students absorb the material at different speeds, but my experience

teaching code has shown me that nearly anyone can master the basics. My goal is to ensure that all students have gained most of what they need to know to begin planning their final projects by mid-semester, if not before.

Final project

You will conduct your final research project on a communication-relevant digital dataset. You will have the opportunity to submit one keyword to me which I will use to generate a dataset for you to use. You'll have your choice of data from Twitter, Facebook, Instagram, TikTok, or your own social media or textual data (pending instructor approval).

The overall goal of your final research project will be to quantitatively describe and explain your dataset using a theoretical framework of your choice. The project will include three distinct quantitative analyses and/or visualizations *which must be thematically linked*. A series of possibilities is listed below; however, these are merely suggestions, and I encourage you to design your own analyses as long as you clear them with me in advance. Here is what you will turn in when you are finished:

- All the code you wrote to generate the analyses
- Any visualizations you created
- A 3,500 - 4,500 word short paper presenting and situating your analyses within a theoretical framework of your choice

You will also give an 12-15-minute oral presentation of your work during our finals period during which you will use the projector to show the visuals you created. Later in the semester I will give instructions and specific examples of what kinds of analyses will count for the final project. Generally, they need to satisfy three criteria:

1. Sufficient complexity (i.e., must require more than one or two lines of code)
2. Can't be something we did in class or one of the homework assignments
3. Worth doing (i.e., you need to be able to explain why the results matter)

I encourage you to be creative and work directly with me to plan analyses that are innovative and interesting.

Grade key

91-100%	H
81-90%	P
71-80%	L
< 70%	F

Required materials

We will install the RStudio environment (which is free) on or before the first day, and we'll use it every class day.

The readings for this class are available on the course Sakai site through the "Course Reserves" link. Please read all assigned readings before the date on which they are listed and come to class prepared to discuss them.

The readings are organized into three units: Concepts, Applications, and Techniques. The *Concepts* readings are intended to give you a high-level introduction to the most important concepts in computational social science. *Applications* focuses on how computational methods have been applied to four of our department's key research areas: news and journalism, political communication, public relations, and health communication. *Techniques* explores the most important logistical issues involved in the computational research process. I will try to connect the *how* track's material to the readings whenever possible.

Due to time constraints in writing the syllabus, the readings are subject to change. I will commit to changing a maximum of one reading per week, and if I do so, I will give you at least a week's notice.

In addition to the seminar readings, I also offer the following supplementary resource to help you with your code. It is freely available online:

- *Stack Overflow:* http://stackoverflow.com/ -- This is an excellent question-and-answer site that almost certainly already has the answer to whatever programming questions you may have. I strongly suggest searching for the answer rather than asking a question yourself—you might get a rather brusque response if you ask a question that's already been asked several times.

My classroom expectations

In this class, I expect that you will:

- Come to class prepared to engage with the day's material
- Come to class on time
- Complete all assignments on time
- Silence your mobile phone during class
- Not waste class time on electronic or online services unrelated to class
- Speak up at least once per class period
- Let me know if and when you're having trouble understanding anything (feel free to do so publicly or privately)
- Not insult or belittle me or your fellow classmates
- Refrain from plagiarism and other violations of UNC's Honor Code (see below)

Additionally, given recent events, I feel it is important to clarify the bounds of class conduct and discussion in advance to reduce confusion about what is permitted and what is not. I undertake

this task in the spirit of one of every university's main purposes: to distinguish between valid and invalid knowledge and judgments. Therefore, over and above UNC's official diversity statement (reproduced below), I hereby establish the following bounds of classroom conduct. All students in this class will:

- Refrain from judging individuals according to the collective groups of which they are members (e.g. race, gender, class, sexual orientation, disability status, etc.);
- Assess intellectual ideas and arguments strictly according to the evidence supporting them, and not based on the identities of the individual who created them;
- Acknowledge that due to historical and contemporary systems of oppression, allegations of racism, sexism, homophobia, ableism, etc. are not symmetrical between social groups. This means such claims can only be valid when advanced by members of a less powerful group against a more powerful group. Allegations in the opposite direction (e.g. of "reverse racism") will not be tolerated. Such notions have been definitively debunked by many strong arguments for which I am happy to provide references upon request.

By the same token, you can expect that I will:

- Come to class prepared and enthused to engage with the day's material
- Treat your personal views with respect
- Carefully explain any concepts that don't make sense
- Cultivate a civil and welcoming class environment
- Return your graded assignments within about a week
- Reward good-faith efforts to engage with course material
- Refer plagiarism and other violations of UNC's Honor Code to the proper authorities (see below)

My policies

- *Lateness and absences:* Please arrive promptly for class; lateness is disruptive and inconsiderate. Chronic lateness will count against your participation grade.
- *Late assignments:* Turning in your assignments on time will be absolutely critical in this class. Otherwise you will fall behind, which will jeopardize your ability to complete the final assignment. So please keep current with these.
- *Mobile phones:* These should not be used during class under any circumstances, and your ringer should be set to silent.
- *Bathroom:* Feel free to use the bathroom whenever you need to; just leave and re-enter as quietly as possible.

University Policies

The Honor Code

It is my duty to report any and all suspected Honor Code violations to the Student Attorney General. If you are not familiar with the Honor Code, please review it at http://instrument.unc.edu . As stated in the Honor Code, "It shall be the responsibility of every student at the University of North Carolina at Chapel Hill to obey and support the enforcement of the Honor Code, which prohibits lying, cheating, or stealing when these actions involve academic process or University student or academic personnel acting in an official capacity."

A special note about plagiarism: The Instrument of Student Governance at UNC defines plagiarism as "deliberate or reckless representation of another's words, thoughts, or ideas as one's own without attribution in connection with submission of academic work, whether graded or otherwise." Copying-and-pasting from online sources without citing the source from which you obtained the content is clearly an instance of plagiarism. However, it may also be plagiarism if you rely too heavily on the structure and reasoning of another piece (for example, if you rely too much on swapping out synonyms or making only very superficial changes to content that is not yours). This type of extensive paraphrasing is not acceptable in this course, which requires you to demonstrate original thinking and analysis. If you have any questions about whether your use of reference material is appropriate, please see me. If any part of your work is judged by me and an independent faculty member to reflect inappropriate use of reference material, I reserve the right to adjust assignment and course grades downwards, in addition to reporting suspected violations as described in the preceding paragraph.

Students with Disabilities

If you have a diagnosed or suspected disability that you think might affect your performance in this course, you should contact Accessibility Resources & Service to determine whether and to what extent services or accommodations are available. If you think this might apply to you, please contact Accessibility Resources & Service at 962-8300 or visit the department's Website at https://accessibility.unc.edu/ . Please understand that I'm not qualified or permitted under University policies to provide any disability-related accommodations without authorization from ARS.

Diversity

The University of North Carolina at Chapel Hill is committed to equality of educational opportunity. The University 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 Dean of Students (Suite 1106, Student Academic Services Building, CB# 5100, 450 Ridge Road, Chapel Hill, NC 27599-5100 or [919] 966-4042) has been designated to handle inquiries regarding the University's nondiscrimination policies.

Course reading schedule

Note: these readings are from the course's previous iteration, and they are subject to change. If they change, I'll send out the new readings no less than one week before they are due.

Homework assignments, which will be assigned most but not all weeks, will be issued on class days and due Tuesdays by noon.

Date	Topic	Readings
		CONCEPTS
1/11	Computational social science basics	Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., … others. (2009). Life in the network: the coming age of computational social science. *Science*, 323(5915), 721. boyd, danah, & Crawford, K. (2012). Critical questions for Big Data. *Information, Communication & Society*, 15(5), 662–679. Shah, D. V., Cappella, J. N., & Neuman, W. R. (2015). Big data, digital media, and computational social science: Possibilities and perils. *The ANNALS of the American Academy of Political and Social Science*, 659(1), 6-13. Lazer, D. M. J., Pentland, A., Watts, D. J., Aral, S., Athey, S., Contractor, N., Freelon, D., Gonzalez-Bailon, S., King, G., Margetts, H., Nelson, A., Salganik, M. J., Strohmaier, M., Vespignani, A., & Wagner, C. (2020). Computational social science: Obstacles and opportunities. *Science, 369*(6507), 1060-1062. https://doi.org/10.1126/science.aaz8170
1/18	Interpreting trace data	boyd, D., Golder, S., & Lotan, G. (2010). Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In 43rd Hawaii International Conference on System Sciences (HICSS) (pp. 1–10). Freelon, D. (in press). Inferring individual-level characteristics from digital trace data: Issues and recommendations. In N. J. Stroud (Ed.). Tufekci, Z. (2014). Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls. In Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media. Ann Arbor, MI: AAAI Publications. Retrieved from http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8062
1/25	Disciplinary differences in computational research	Adamic, L. A., & Glance, N. (2005). The political blogosphere and the 2004 US election: divided they blog. In Proceedings of the 3rd international workshop on Link discovery (pp. 36–43). New York, NY: ACM. Freelon, D. (2015). On the cutting edge of big data: Digital politics research in the social computing literature. In S. Coleman & D. Freelon (Eds.), *Handbook of Digital Politics* (pp. 451–472). Northampton, MA: Edward Elgar. Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559
2/1	Software development	Araujo, T., Lock, I., & Velde, B. van de. (2020). Automated Visual Content Analysis (AVCA) in Communication Research: A Protocol for Large Scale Image Classification with Pre-Trained Computer Vision Models. *Communication Methods and Measures, 14*(4), 239–265. https://doi.org/10.1080/19312458.2020.1810648 Freelon, D. (2018). Partition-specific network analysis of digital trace data: Research questions and tools. In S. Gonzalez-Bailon & B. F. Welles (Eds.), *Oxford Handbook of Networked Communication*. New York: Oxford University Press. Thelwall, M. (2017). Heart and Soul: Sentiment Strength Detection in the Social Web with SentiStrength. In J. Hoylst (Ed.), *Cyberemotions* (pp. 119–134). Springer.

2/8	Computational ethics	Metcalf, J., & Crawford, K. (2016). Where are human subjects in Big Data research? The emerging ethics divide. *Big Data & Society, 3*(1), 2053951716650211. https://doi.org/10.1177/2053951716650211 Zimmer, M. (2010). "But the data is already public": on the ethics of research in Facebook. *Ethics and Information Technology,* 12(4), 313–325. Zook, M., Barocas, S., Boyd, D., Crawford, K., Keller, E., Gangadharan, S. P., … Pasquale, F. (2017). Ten simple rules for responsible big data research. *PLOS Computational Biology*, 13(3), e1005399.

APPLICATIONS

2/15	News/journalism	Hermida, A., Lewis, S. C., & Zamith, R. (2014). Sourcing the Arab Spring: A Case Study of Andy Carvin's Sources on Twitter during the Tunisian and Egyptian Revolutions. *Journal of Computer-Mediated Communication, 19*(3), 479–499. https://doi.org/10.1111/jcc4.12074 Jacobi, C., Atteveldt, W. van, & Welbers, K. (2016). Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism, 4*(1), 89–106. https://doi.org/10.1080/21670811.2015.1093271 Neuman, W. R., Guggenheim, L., Jang, S. M., & Bae, S. Y. (2014). The Dynamics of Public Attention: Agenda-Setting Theory Meets Big Data. *Journal of Communication*, 64(2), 193–214.
2/22	Politics/ social movements	Bastos, M. T., & Mercea, D. (2016). Serial activists: Political Twitter beyond influentials and the twittertariat. *New Media & Society*, 18(10), 2359–2378. Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science*, 26(10), 1531–1542. Freelon, D., McIlwain, C., & Clark, M. D. (in press). Quantifying the power and consequences of social media protest. *New Media & Society.*
3/1	Strategic communication	Fröhlich, R., & Jungblut, M. (2018). Between factoids and facts: The application of 'evidence' in NGO strategic communication on war and armed conflict. *Media, War & Conflict, 11*(1), 85–106. https://doi.org/10.1177/1750635217727308 Guo, C., & Saxton, G. D. (2017). Speaking and Being Heard: How Nonprofit Advocacy Organizations Gain Attention on Social Media. *Nonprofit and Voluntary Sector Quarterly.* Hagen, L., Keller, T., Neely, S., DePaula, N., & Robert-Cooperman, C. (2017). Crisis Communications in the Age of Social Media: A Network Analysis of Zika-Related Tweets. *Social Science Computer Review*, 0894439317721985. https://doi.org/10.1177/0894439317721985
3/8	Health communication	Eichstaedt, J. C., Schwartz, H. A., Kern, M. L., Park, G., Labarthe, D. R., Merchant, R. M., … Seligman, M. E. P. (2015). Psychological Language on Twitter Predicts County-Level Heart Disease Mortality. *Psychological Science, 26*(2), 159–169. https://doi.org/10.1177/0956797614557867 Jashinsky, J., Burton, S. H., Hanson, C. L., West, J., Giraud-Carrier, C., Barnes, M. D., & Argyle, T. (2014). Tracking Suicide Risk Factors Through Twitter in the US. *Crisis*, 35(1), 51–59. Tempel, J. van der, Noormohamed, A., Schwartz, R., Norman, C., Malas, M., & Zawertailo, L. (2016). Vape, quit, tweet? Electronic cigarettes and smoking cessation on Twitter. *International Journal of Public Health*, 61(2), 249–256.
3/15	*Spring break…*	

TECHNIQUES

3/22	Collecting data	Driscoll, K., & Walker, S. (2014). Working Within a Black Box: Transparency in the Collection and Production of Big Twitter Data. *International Journal of Communication*, 8(0), 20. Freelon, D. (forthcoming). When social media data disappear. In E. Hargittai (Ed.), *Research Confidential III.* New York. Columbia University Press. González-Bailón, S., Wang, N., Rivero, A., Borge-Holthoefer, J., & Moreno, Y. (2014). Assessing the bias in samples of large online networks. *Social Networks*, 38, 16–27. https://doi.org/10.1016/j.socnet.2014.01.004
3/29	Cleaning data/ information quality	Blank, G. (2017). The Digital Divide Among Twitter Users and Its Implications for Social Research. *Social Science Computer Review,* 35(6), 679–697. Denny, M. J., & Spirling, A. (2018). Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it. *Political Analysis, 26*(2), 168-189. Zahedi, Z., & Costas, R. (2018). General discussion of data quality challenges in social media metrics: Extensive comparison of four major altmetric data aggregators. PLOS ONE, 13(5), e0197326. https://doi.org/10.1371/journal.pone.0197326
4/5	Statistical considerations	Frické, M. (2015). Big data and its epistemology. *Journal of the Association for Information Science and Technology*, 66(4), 651–661. https://doi.org/10.1002/asi.23212 Hoerl, R. W., Snee, R. D., & De Veaux, R. D. (2014). Applying statistical thinking to 'Big Data' problems. *Wiley Interdisciplinary Reviews: Computational Statistics*, 6(4), 222–232. Shmueli, G. (2010). To Explain or to Predict? *Statistical Science*, 25(3), 289–310.
4/12	Automated text analysis 1	Grimmer, J., & Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, 21(3), 267–297. Welbers, K., Atteveldt, W. V., & Benoit, K. (2017). Text Analysis in R. *Communication Methods and Measures, 11*(4), 245–265. https://doi.org/10.1080/19312458.2017.1387238 Zamith, R., & Lewis, S. C. (2015). Content Analysis and the Algorithmic Coder: What Computational Social Science Means for Traditional Modes of Media Analysis. *The ANNALS of the American Academy of Political and Social Science,* 659(1), 307–318.
4/19	Automated text analysis 2	Barberá, P., Boydstun, A. E., Linn, S., McMahon, R., & Nagler, J. (2021). Automated Text Classification of News Articles: A Practical Guide. *Political Analysis, 29*(1), 19–42. https://doi.org/10.1017/pan.2020.8 Charlesworth, T. E. S., Yang, V., Mann, T. C., Kurdi, B., & Banaji, M. R. (2021). Gender Stereotypes in Natural Language: Word Embeddings Show Robust Consistency Across Child and Adult Language Corpora of More Than 65 Million Words. *Psychological Science, 32*(2), 218–240. https://doi.org/10.1177/0956797620963619 Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri, H., & Adam, S. (2018). Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology. *Communication Methods and Measures, 12*(2–3), 93–118. https://doi.org/10.1080/19312458.2018.1430754
4/26	Network analysis	Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network Analysis in the Social Sciences. *Science*, 323(5916), 892–895. Can, U., & Alatas, B. (2019). A new direction in social network analysis: Online

social network analysis problems and applications. *Physica A: Statistical Mechanics and Its Applications, 535*, 122372. https://doi.org/10.1016/j.physa.2019.122372

Himelboim, I. (2017). Social Network Analysis (Social Media). *The International Encyclopedia of Communication Research Methods*, 1–15.

5/1 (Mon)	The exam period runs from **8am to 11am,** and the final project is due at 8am via Sakai**.**