AIT 602: Introduction to Research in Applied IT
(Research Methods in Information Science)

Spring 2022

When: Tuesdays 16:30 PM – 19:10 PM
Where: Hybrid
Offline Sessions: Thompson Hall 1017
Online Sessions: Zoom link will be disseminated through BlackBoard & Calendar Invite
Instructor: Dr. Myeong Lee (mlee89@gmu.edu)
Instructor Office hours: By appointment
TA: Rafael Hung-Mao Chen
TA Office hours: TBA
Prerequisites: None (no prior experience in research is required)

Goal: This course aims to help students learn about research-driven thinking in building technological tools and analyzing data. Data analytics work requires thorough understanding of the nature of data, causality/correlation concerns, and justifications on data analysis methods. This course is designed to cover these aspects through discussing research-driven thinking processes, methodologies, theoretical concerns, tools/skills, and methods that need to be considered in analyzing data or building systems. By understanding various tools, methods, and philosophies behind data science work, this course aims to help students understand why and how analysis and development strategies can be designed for answering research questions. An ultimate goal of this course is to help students independently develop research projects and design data analytics strategies.

Midterm: Open book tests with time limit.
Final exam: Group project and presentation
Course requirements and evaluations:

- Class attendance and participation (10%)
- Homework assignments (35%)
- Midterm Proposal (20%)
- Final Project (25%)
- Two Quizzes (10%)

Recommended course books and resources:

- Picard, A. (2013). *Research Methods in Information*. ALA.
- Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process*. SAGE.
- Jennifer Golbeck (2015). *Introduction to Social Media Investigation: A Hands-on Approach*. Online access through GMU: http://proquest.safaribooksonline.com/book/web-applications-and-services/social-media/9780128016565

- Shamanth Kumar, Fred Morstatter, Huan Liu (2014). *Twitter data analytics*: Springer Online access through GMU: https://link.springer.com/book/10.1007/978-1-4614-9372-3
- Course materials include research papers and web resources in each session.

Course lecture structure:

- Group discussion on a particular topic (1) in class and (2) on Blackboard
- Short tutorials on "how to do something"
- Hands-on experience in class
- Homework assignment on what you have learned in class
- Mid-term and final presentations on research project

Student Outcomes and Expectations

- Upon completion of the course, students are able to design and review a research study independently.
- Upon completion of the course, students are able to design data analysis strategies in the way that answers particular research questions.
- Upon completion of the course, students understand why and how certain data modeling or analysis methods are designed among other methods.
- During the course, students are expected to actively participate in in-class and online discussions.
- During the course, students are expected to read the assigned reading materials before each session.

Weekly Schedule (Course Structure)

Session	Course Materials and Resources
Week 1 Foundations of Applied IT and Information Science Research	**Main Topics:** • Research process (inductive and deductive), epistemology and ontology (what is knowing and how does one know), course logistics • Project team formation **Reading List:** • Crotty Ch. 1, 2, 3 • Git Tutorial: https://github.com/myeong/git-practice **Class Activity:** • Git Process **Assignment 1:** • Finish the human-subject research module on the CITI website (IRB) **Online Discussion 1:** (1) Introduce yourself on the discussion board (2) Discuss a data analysis topic of your interest from epistemological, theoretical, and methodological perspectives.
Week 2 Literature Review, Research Process, and Social Media	**Main Topics:** • Basics in research methods, literature review, and social media in research **Reading List:**

	• Research process: Link • Background/Literature review: Cronin, P., Ryan, F., & Coughlan, M. (2008). Undertaking a literature review: a step-by-step approach. British journal of nursing (Mark Allen Publishing) (17), 38-43. Link • Research in the Crowdsourcing Age, a Case Study. Linkc **Supplemental Reading:** • Pickard, A. J. (2013). Chapter 2: Reviewing literature. In *Research Methods in Information* (2nd ed., pp. 25-38). Chicago, IL: Neal-Schuman. • Krathwohl, D. R. (2009). Chapter 5: Creating research problems. In *Methods of Educational and Social Science Research* (3rd ed., pp. 83-104). Long Grove, IL: Waveland Press, Inc. • Hernon, P. & Schwartz, C. (2007). What is a problem statement? *Library & Information Science Research, 29*(3), 307-309. **Class Activity:** • Inspecting Twitter data analysis provided by an automatic analysis app: Account Analysis App **Online Discussion 2:** • Discuss the pros and cons of using crowdsourcing techniques (e.g., Amazon M-Turk) and their relationships to epistemology, after reading "Research in the Crowdsourcing Age." Providing a good analysis would be a plus.
Week 3 Quantitative Methods I: Data Curation, Cleaning, and Crowdsourcing	**Main Topics:** • The history of data curation practices. • Computational power-assisted data curation methods. • Presentation by Theresa Calcagno, IT & Engineering Librarian at VSE. **Reading List:** • Quinn, A. J., & Bederson, B. B. (2011). Human computation: a survey and taxonomy of a growing field., ACM *CHI*. • Wiggins, A., & Crowston, K. (2011). From conservation to crowdsourcing: A typology of citizen science., *IEEE HICSS* 2009 • Karasti, H., Baker, K. S., & Halkola, E. (2006). Enriching the notion of data curation in e-science: data managing and information infrastructuring in the long term ecological research (LTER) network. *Computer Supported Cooperative Work (CSCW), 15*(4), 321-358. • Ruths, D., & Pfeffer, J. (2014). Social media for large studies of behavior. *Science, 346*(6213), 1063-1064. Link **Supplemental Resources:** • IT Library Research guide: https://infoguides.gmu.edu/IST • Google Forms Full Tutorial From Start To Finish - How To Use Google Forms

	• TED Talk: The voices of Twitter users - Evan Williams **Class Activity:** • Amazon M-Turk. To prepare for the class activity, create an account and get approved before you come to the class (both as worker and requester). **Assignment 2:** • Select a research topic on social computing and write a 1000-word literature review. **Online Discussion 3:** • Select one research study that used crowdsourced data or social media data, briefly explain what the study did, and discuss (1) the limitations of the study in answering their questions and (2) how they could improve their study.
Week 4 Quantitative Methods II: Survey and Sampling	**Main Topics:** • Survey methodology and sampling **Reading List:** • Pickard (2013). Ch 5 and 9. • Groves et al., (2004). Ch. 7: Questions and Answers in Surveys • Antin, J., & Shaw, A. (2012). Social desirability bias and self-reports of motivation: a study of Amazon Mechanical Turk in the US and India. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2925-2934). • Blumenstock, J., Cadamuro, G., & On, R. (2015). Predicting poverty and wealth from mobile phone metadata. *Science*, *350*(6264), 1073-1076. **Supplemental Resources:** • Likert scale options for surveys **Class Activity:** • Amazon M-Turk data analysis using R (i.e., crowdsourced survey data analysis) **Online Discussion 4:** • Select one research study that used a survey/sampling method in computational research, briefly explain what the study did, and discuss (1) the limitations of the study in answering their questions and (2) how they could improve their study.
Week 5 Quantitative Methods III: Experimental Design + Quiz 1	**Main Topics:** • Experimental design and computational research that use experiments. **Reading List:** • Pickard (2013). Ch 10. • Bail, C. A., et al. (2018). Exposure to opposing views on social media can increase political polarization. *PNAS, 115*(37), 9216-9221. (available from the course contents)

	• Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *PNAS, 111*(24), 8788-8790. (available from the course contents) • Brennan, K., & Resnick, M. (2012). New frameworks for studying and assessing the development of computational thinking. In *Proceedings of the 2012 annual meeting of the American Educational Research Association*, Vancouver, Canada. Link **Supplemental Resources:** • TED Talk: Simplicity Sells – David Pogue • Experimental Research & Usability evaluation methods: Different methods for evaluating Usability **Class Activity:** • Experimental data analysis (Ref: Link). **Online Discussion 5:** • Discuss the ethical concerns and potential issues in a large-scale social experiment. Read Kramer et al. (2014) before discussing this. Also, discuss any solutions to ethically conduct similar data science research using experimental methods.
Week 6 Quantitative Methods IV: Statistical Methods + Review and Applications	**Main Topics:** • Statistical methods and computational adjustments. • Review of the first half of the course. **Reading List:** • Pickard (2013). Ch 24. • Wang, P., Lee, M., Meng, X., & Butler, B. (2016). Toward an ecology theory of creativity in IT products: A study of mobile device industry. In *International Conference on Information Systems (ICIS) Proceedings*. • Lee, E., Lee, J.-A., Moon, J. H., & Sung, Y. (2015). Pictures speak louder than words: Motivations for using Instagram. *Cyberpsychology, Behavior, and Social Networking, 18*(9), 552-556. • Mull, I. R., & Lee, S.-E. (2014). "PIN" pointing the motivational dimensions behind Pinterest. *Computers in Human Behavior, 33*, 192-200. **Supplemental Resources:** • How to setup TAGS (step by step guide) • Any time you want to download a particular hashtag's data-set follow these instructions • Twitter data collection code **Class Activity:** • Twitter data analysis using statistical methods in R. **Online Discussion 6:**

	• Find a statistical analysis case (either research or practical analysis) and discuss whether the authors tested all the assumptions properly. If so, or if not, discuss how they could improve the original study.
Week 7 Mid-term Presentation	**Main Topics:** • Mid-term presentation on a research proposal (20 minutes per team + 10 minutes Q&A). • Mid-term report needs to be submitted separately. **Evaluation Criteria:** • Did the presentation provide a good justification for "what the research topic is and why it is important"? (20%) • Did the presentation provide a good synthesis of the literature, beyond a simple summary of them? (20%) • Did the presentation derive "hard/good" research questions out of the literature review in a logical way? (20%) • Did the presentation provide reasonable data analysis plans that are effective in answering your research questions? (20%) • Was the presentation overall well organized and provide good future plans and expected outcomes/deliverables? (20%) • [Optionally] Provide and justify budget, if any.
Week 8 Qualitative Methods I: Interview, Focus Group, and Ethnography in Information/Computing Fields	**Main Topics:** • Overview of qualitative research methods. • Ethnography and interviews. • Application of qualitative methods in data science research. **Reading List:** • Qu, S. Q., & Dumay, J. (2011). The qualitative research interview. *Qualitative research in accounting & management*. (available from the course contents) • *Focus Groups*. Qualitative research guideline groups. Retrieved: http://www.qualres.org/HomeFocu-3647.html • Dourish, P. (2008). What ethnography can teach designers (video). Link • Cranshaw, J., Schwarsz, R., Hong, J., & Sadeh, N. (2012, May). The Livehoods project: Utilizing social media to understand the dynamics of a city. In *Sixth International AAAI Conference on Weblogs and Social Media*. Link **Supplemental Resources:** • Johri, A. (2014). *Conducting Interpretive Research in Engineering Education* (chapter contains tips on how to do qualitative research including protocols): Chapter 28.pdf **Class Activity:** • Small-scale interview coding (manuscript given in class) **Online Discussion 7:**

	• In computational research, why is it sometimes necessary to use qualitative methods such as interviews and ethnography? Explain it with any examples.
Week 9 Content Analysis and NLP for Information/Computing Research	**Main Topics:** • Qualitative content analysis method. • LIWC software package for scalable text analysis • Natural Language Processing (NLP) methods. **Reading List:** • Mayring (2000). Qualitative Content Analysis. *Forum: Qualitative Social Research* (available from the course contents) • Bender, J. L., Jimenez-Marroquin, M. C., & Jadad, A. R. (2011). Seeking support on Facebook: a content analysis of breast cancer groups. *Journal of medical Internet research, 13*(1), e16. (available from the course contents) • Tausczik, Y., & Pennebaker, J.W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, 29*, 24-54. (available from the course contents) **Supplemental Resources:** • Kurian, J. C., & John, B. M. (2017). User-generated content on the Facebook page of an emergency management agency: A thematic analysis. *Online Information Review, 41*(4), 558-579. • Golbeck, J., Grimes, J. M., & Rogers, A. (2010). Twitter use by the US Congress. *Journal of the Association for Information Science and Technology, 61*(8), 1612-1621. • Hu, Y., Manikonda, L., & Kambhampati, S. (2014). *What We Instagram: A First Analysis of Instagram Photo Content and User Types.* Paper presented at the ICWSM. **Class Activity:** • Qualitative Twitter content analysis for NLP. **Assignment 3:** • Qualitative content analysis of Twitter data about COVID-19 and make a short report out of it. Using any theoretical framework or word classifiers, manually code the sample Twitter dataset (~100 tweets) and find the patterns that describe the target data stream. These patterns do not need to be found computationally, but qualitatively. Report your findings descriptively (up to 2 pages). **Online Discussion 8:** • Watch a video -- Dourish, P. (2008). What ethnography can teach designers (video). Link and discuss why ethnographic work could be useful in computational research.
Week 10 Social Network Analysis, Computational	**Main Topics:** • Quiz 2 • Social network analysis and research streams on network/graph.

Modeling, and Geospaital Data Analysis + Quiz 2	• Computational modeling and geospatial data analysis models. **Reading List:** • Aggarwal, C. C. (2011). *An introduction to social network data analytics. In Social network data analytics* (pp. 1-15). Springer, Boston, MA. (available from the course contents) • Castillo, C., Mendoza, M., & Poblete, B. (2011, March). Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web* (pp. 675-684). (available from the course contents) • What is Spatial Data - An Introduction to Spatial Data and its Applications (YouTube Video by Fullstack Academy) Link **Supplemental Resources:** • McKenzie, G., Liu, Z., Hu, Y., & Lee, M. (2018). Identifying urban neighborhood names through user-contributed online property listings. *ISPRS International Journal of Geo-Information, 7*(10), 388. • First Evidence That Social Bots Play a Major Role in Spreading Fake News • The hidden influence of social networks-TED TALK • Classifying Twitter Topic Networks using Social Network Analysis **Class Activity:** • Twitter network analysis using R. **Online Discussion 9:** • Survey social network analysis and visualization tools (at least 5 different tools) and discuss the pros and cons of each tool.
Week 11 Visualization and System Design Methodologies	**Main Topics:** • Visualization methods, tools, and techniques. • System design methodologies. **Reading List:** • Rosson, M. B., & Carroll, J. M. (2009). Scenario-based design. In *Human-computer interaction* (pp. 161-180). CRC Press. (available from the course contents) • Segel, E., & Heer, J. (2010). Narrative visualization: Telling stories with data. *IEEE transactions on visualization and computer graphics, 16*(6), 1139-1148. (available from the course contents) • Muller, M. J. (2007). Participatory design: the third space in HCI. In *The human-computer interaction handbook* (pp. 1087-1108). CRC press. (Read for pages 1-16. Available from the course contents) **Supplemental Resources:** • Visualization Tour: https://queue.acm.org/detail.cfm?id=1805128 • D3: d3js.org • TED talk: The beauty of data visualization

	Class Activity: • Twitter data visualization using Plot.ly (JS package) **Assignment 4:** • Social network analysis using Twitter dataset. You can download a sample Twitter dataset, i.e., follower-followee network data from the course Github. Load this data to a social network analysis tool such as NodeXL or Gephi. Then, calculate the degree centrality, betweenness centrality, and closeness centrality measures for each node and find the top-5 nodes for each measure. Also, visualize the network. Then, write a 1-page report that discusses what the results mean in terms of social network structure. **Online Discussion 10:** • Find two visualization techniques and discuss the pros and cons of the two techniques. For what kind of data would each of the visualization techniques be more effective and why? Discuss it with example cases.
Week 12 Theory-driven Research with Data: Grounded Theory, Social Theories, and their Applications	**Main Topics:** • What is theory? • Theory development process and computational research based on theories. **Reading List:** • Sutton, R. I., & Staw, B. M. (1995). What theory is not. *Administrative science quarterly*, 371-384. (Available from the course contents) • Charmaz, K., & Belgrave, L. (2012). Qualitative interviewing and grounded theory analysis. *The SAGE handbook of interview research: The complexity of the craft*, *2*, 347-365. (available from the course contents) • Winkelman, W. J., Leonard, K. J., & Rossos, P. G. (2005). Patient-perceived usefulness of online electronic medical records: employing grounded theory in the development of information and communication technologies for use by patients living with chronic illness. *Journal of the American Medical Informatics Association*, *12*(3), 306-314. (available from the course contents) • Aral, S., & Walker, D. (2014). Tie strength, embeddedness, and social influence: A large-scale networked experiment. *Management Science*, *60*(6), 1352-1370. (available from the course contents) **Class Activity:** • Quantifying diversity measures using Census data. **Online Discussion 11:** • Doing "science" usually involves theory-development processes. However, it is not always the case that computational researchers develop research questions based on theoretical framework; they

	also develop questions out of data without theories. (1) Disucss the pros and cons of these two different approaches and (2) provide your preferences when developing research questions in the field of Information Science (and why).
Week 13 Research Progress Check, Course Review, and Q&A	**Main Topics:** • Research Progress Check and Course Review • Q&A for your Final Projects
Week 14 Final Presentation	**Main Topics:** • Final-term presentation on a completed research (20 minutes per team + 10 minutes Q&A). • Final research report needs to be submitted separately. **Evaluation Criteria:** • Did the presentation provide a good justification for "what the research topic is and why it is important"? (20%) • Did the presentation provide a good synthesis of the literature, beyond a simple summary of them and derived a good/hard question out of it? (20%) • Did the study provide reasonable data analyses that are effective in answering the research questions? (20%) • Were the data analysis results well discussed in the way that provide implications for other researchers in the field as well as practitioners? (20%) • Was the presentation overall well organized and provide reasonable explanations on the limitations and future work? (20%)

Honor Code

All work performed in this course will be subject to GMU's Honor Code. Students are expected to do their own work in the course unless a group project is approved by the instructor. In papers and project reports, students are expected to write in their own words, rather than cutting-and-pasting from sources found on the Internet. The goal of assignments is to demonstrate what you have learned, not what you can google. When you do use text or graphical material from books, articles, and the Web, enclose the material in quotes and provide a complete and proper reference (in APA format). If a paragraph is used then it should be indented in the text (both left and right margins). In-text citation can use the [Author, Year] format or the Numerical [1] format which must refer to the source in the References section of your assignment. Use APA for guidance on citation style, usage, etc. (Don't buy the big CMS. See the smaller A Manual for Writers by Kate Turabian). Regardless of the citation method used, proper citations always include: Author(s), Title, Publication Date, Publisher, and URL (if from the Web, along with Last Accessed Date). BlackBoard's SafeAssign service will be used to review selected student assignments. The followings are additional honor code items:

- Wikipedia is **not** a primary reference. Use it for initial discovery, but use and cite primary references (which Wikipedia itself might use).
- If you need assistance with writing an assignment, you can get assistance here: http://writingcenter.gmu.edu
- Refer the Graduate Policies for general policies about courses and degrees: https://catalog.gmu.edu/policies/academic/graduate-policies/
- Any programming/coding assignments must adhere to the CS Honor Code.

Notes

- The course will be structured based on the assumption that you have read papers.
- Lecture slides from instructor's material will be posted on Blackboard.
- E-mail the instructor if you anticipate being unable to meet any course requirements in a timely manner.
- Personal Safety and Security: The Mason Alert system provides emergency information of various sorts. Students can sign up for it by visiting the website https://alert.gmu.edu. Students are also reminded that an emergency poster exists in each classroom explaining what to do in the event of crises and that further information about emergency procedures exists on https://ready.gmu.edu/be-prepared/
- Computer and IT Security: Visit GMU's IT http://itsecurity.gmu.edu/ web site regularly. Norton AntiVirus Software is free to download for all GMU students/faculty/staff.

Important Dates

Dates for adding, dropping the course, etc. are available via: https://registrar.gmu.edu

Religious Holidays

A list of religious holidays is available on the GMU's Religious Holiday Calendar. GMU respects any religious holidays. However, any student whose religious observance conflicts with a scheduled course activity must contact the Instructor at least 2 modules in advance of the conflict date in order to make alternative arrangements.

Attendance Policy

Scheduled course sessions will be spent on clarification, amplification, and review of material through the use of slides, examples, and exercises. Lecture slides are complements to the lecture session, not substitutes for it. Each course session is an excellent time for you to raise questions, request additional examples, and get explanations of any ideas that are still unclear to you. As members of the academic community, all students are expected to contribute regardless of their proficiency with the subject matter. Students are expected to make prior arrangements with Instructor if they know in advance that they will miss any class and to consult with the Instructor as soon as possible if they miss any class without prior notice. Any student who expects to miss more than one class session is advised to drop the course and take it in a later semester when he/she can attend every class.

Privacy

The Instructor will not discuss issues relating to an individual student with anyone lacking a need to know without prior written permission of the student. This includes a student's family members and other students. Under no circumstances will a student's graded work be returned to another student. Instructors, staff, and Teaching Assistants will take care to protect the privacy of each student's scores and grades.

Disability Accommodations

The Office of Disability Services (ODS) works with students with disabilities to arrange for appropriate accommodations to ensure equal access to university services. Any student with a disability of any kind is strongly encouraged to register with ODS as soon as possible and take advantage of the services offered. Phone: 703-993-2474, Web: http://es.gmu.eeu.

Writing Center

A114 Robinson Hall; (703) 993-1200; http://writingcenter.gmu.eeu

Counseling and Psychological Services (CAPS)

(703) 993-2380 http://caps.gmu.eeu