Behavioral Data Science

PSYC UN1930 (4 points)

Instructor: Matthew R. Sisco

ms4403@columbia.edu

Office: Schermerhorn 403C

Office hours: Mondays 4-5pm and available upon request

Fall 2020

Mondays, 2:10 – 4:00pm

Location: Zoom

Syllabus

Bulletin Description: This course covers the basic skills and knowledge needed to address psychological research questions using data science methods. Topics cover the full scope of a behavioral data science research project including data acquisition, data processing, and data analysis.

Course Description

This course covers the fundamental skills and knowledge needed for using data science methods to answer research questions in psychology. Topics will cover the full scope of skills and knowledge needed to complete a basic behavioral data science project including data acquisition (e.g. collecting data through APIs and web scraping), data processing (e.g. high performance computing and feature engineering), and data analysis (including machine learning, natural language processing, and advanced regression analyses). Discussion papers from the expanding literature at the intersection of behavioral science and data science will be examined and discussed. The papers will provide concrete examples of the techniques taught and will show the breadth of possible research designs. We will focus on papers addressing psychological research questions, and will also evaluate some relevant papers from other disciplines studying human behavior. The coursework will involve both empirical and methodological readings, and a series of skills-focused lab assignments (in the R programming language). As a core component of the class, students will design and implement their own behavioral data science research projects using methods taught in the course.

Prerequisites

- Science of psychology (PSYC UN1001) or similar

- AND Introduction to Statistics (PSYC UN1610 or equivalent) (you should be comfortable with simple linear regression models)
- AND Research methods (PSYC UN14xx or equivalent) or lab experience with research methods
- AND novice to intermediate R programming experience (you should be familiar with 80% or more of the basic R concepts and functions listed here: https://rstudio.com/wp-content/uploads/2016/10/r-cheat-sheet-3.pdf)
- AND Instructor permission

Course Objectives

Through completion of this course, you will:

a) Gain a broad theoretical and practical understanding of data science methods applicable to behavioral research

b) Critically evaluate current approaches, research methods, and empirical papers in the field

c) Constructively discuss relevant literature in class

d) Develop your research communication skills, both oral and written

e) Draw on course content to develop your own original research question and research proposal

f) Design and implement a research project empirically evaluating your research question

Course Role in Departmental Curriculum

This course is suitable for advanced undergraduates and postbacs majoring or concentrating in Psychology. Students pursuing a major in Neuroscience and Behavior will also receive registration priority. Students majoring in Economics, Sociology, or Political Science are eligible to enroll, however students majoring/concentrating in Psychology or Neuroscience and Behavior will be given registration priority.

Completion of this course satisfies the following psychology department requirement:

• Additional psychology elective course

Note: this course does not fulfill the psychology seminar requirement

Course Grading & Requirements

10%	1. Class participation (5% attendance, 5% class discussion participation)
10%	2. Thought papers
20%	3. Weekly lab assignments (10)
15%	4. Project proposal (5% presentation, 10% written proposal)
35%	5. Final project (10% presentation, 25% written paper)

1. Class participation: 10%

You are expected to attend and actively participate in every class. You should not only share your own thoughts on the discussion readings throughout the class, but also raise questions encouraging your peers to share theirs. Additionally, you will be expected to give your peers constructive feedback on their research proposals. Your participation will be evaluated after every class – as such, you will be penalized for any unexcused absences. Feel free to come see me anytime throughout the course to ask for feedback or suggestions regarding your class participation (or of course, to further discuss an idea that was raised in class). Participating in class can be more difficult for some students, and if that's the case, I encourage you to come see me at the beginning of the semester so that we can work out ways you can contribute. In these cases, later participation will be weighed more heavily to reward improvement.

2. Thought papers: 10%

By 11PM the night before each class, you are required to submit a short thought paper on Canvas (roughly 150-250 words in length). The goal of these thought papers is to promote active reading and critical thinking, and to stimulate thoughts to discuss in class: you can raise theoretical or methodological questions related to the readings, share insights or comment on the implications of empirical findings, or relate the readings to previous class discussions. Try to integrate two or more readings into each thought paper. Be prepared to share your thoughts with your peers. These will not be formally graded but will be checked for completion/effort (each worth 1 point [those completed but with a clear lack of effort will receive half credit – note that greater length does not necessarily indicate greater effort]). Students can miss one thought paper during the semester at no penalty (but 1 extra credit point will be added to your final grade in the course if you complete all 11 of them).

3. Weekly lab assignments: 20%

After each class meeting, students will have the remainder of the week to complete the related lab assignment. The lab assignments are due (submitted via Canvas) at the start of class the following week. There will be 10 graded assignments worth 20 points in total. Students will receive 2 points for each complete, on-time assignment, 1.5 points for each late assignment, and no points for incomplete assignments or assignments later than two weeks after the due date. The weekly lab assignments allow students to implement the techniques and methods discussed in class. There is some flexibility in the data and specifics of analyses students can use in the lab assignments. It is highly recommended that students use the lab assignments to prepare for the final project. It is perfectly acceptable to use analyses run in the lab assignments in the final project.

4. Project proposal: 15% (5% presentation, 10% written proposal)

Mid-way through the semester, students will submit a (500-1,000 words) project proposal and present a five minute summary of their proposed research project to the class. Each student is required to schedule a meeting with me to discuss their ideas for a project proposal by the week before the day the proposals and presentations are due. At the end of the semester, students will summarize the analyses of their own data in a 10-15 minute class presentation and written paper based on the project proposed mid-way through the semester (more details on this final paper below).

Proposal presentation rubric:

Clear research question(s): 1pt

Discusses connection with past research: 1pt

Clear plan(s) for data collection: 1pt

At least general plan for analysis: 1pt

Discusses implications of possible results: 1pt

Proposal structure:

a. Introduction: Research question & brief literature review

b. Proposed Data: Source of data and planned extraction procedures

c. Proposed Method: Variables of interest and analysis procedure

d. Predicted Results: Description of anticipated results (feel free to visualize)

e. Discussion: Implications and limitations of predicted results

f. References: 5+ references

5. Final project: 35% (10% presentation, 25% written paper)

The research paper should (loosely) follow APA format with a brief introduction to the topic, a detailed methods section, a thorough results section, and a concise discussion. The paper should be between 1,500 and 3,000 words (not including references, figure captions, or tables). Final papers are due by submitting on canvas or paper copy on December 21st at midnight and are worth 25 points.

Your papers will be graded based on thoroughness of the literature review (20%), integration of relevant and empirically valid methodology (20%), reasoning and implementation of the chosen analysis (25%), thoughtfulness of discussion (20%), overall presentation (grammar, spelling, APA formatting, etc.) (5%), and creativity and originality of the proposed idea (10%). The last grading criterion, creativity and originality, can be met by synthesizing ideas or approaches discussed in class to design your project, rather than simply replicating a discussed design with a minor variation.

In addition, you will present your research proposal (approximately 10-12 minutes including time for questions) to your classmates on the last day of class which is worth 10 points. I will discuss these presentations in more detail throughout the term.

Course Policies

Attendance:

Absences will be excused with the presentation of proper documentation (i.e. doctor's or dean's note). Please inform me of the absence as soon as possible. You will still be responsible for completing the work due that particular class session. Each unexcused absence will result in 0.8 (1/13 * 10) points deducted from your grade.

Late work: Unless excused by a Doctor's or Dean's note:

• Thought papers: Given that the purpose of thought papers is to prepare for the class discussion, you cannot submit a thought paper after class. Some leniency will be afforded for timing: half of your grade (0.5 points) will be deducted past the 11 PM deadline as long as it is submitted before the start of class.

• Lab assignments: 0.5 points off for late lab assignments.

• Project proposal: 1 point of your grade for the project proposal paper will be deducted per day the proposal is late. (Reminder the project proposal is worth 15 points total including the paper and presentation.)

• Final project: 1 point of your grade for the final project paper will be deducted per day the paper is late. (Reminder the final project is worth 35 points total including the paper and presentation.)

Class Etiquette:

Cell phones are not allowed to be taken out in class and should be kept on silent (not vibrate). Laptops or tablets may be used for anything course related. However, out of courtesy to your classmates and respect for your own learning, please refrain from using laptops or tablets for any other purpose.

Students with Disabilities:

If you are a student with a disability and have a DS-certified 'Accommodation Letter' please come to my office hours by the end of Week 2 to confirm your accommodation needs. If you believe that you might have a disability that requires accommodation, you should contact Disability Services at 212-854-2388 and disability@columbia.edu.

Academic Integrity:

Columbia University Undergraduate Guide to Academic Integrity:
http://www.college.columbia.edu/academics/academicintegrity

Faculty Statement on Academic Integrity:

The intellectual venture in which we are all engaged requires of faculty and students alike the highest level of personal and academic integrity. As members of an academic community, each one of us bears the responsibility to participate in scholarly discourse and research in a manner characterized by intellectual honesty and scholarly integrity. Scholarship, by its very nature, is an iterative process, with ideas and insights building one upon the other. Collaborative scholarship requires the study of other scholars' work, the free discussion of such work, and the explicit acknowledgement of those ideas in any work that inform our own. This exchange of ideas relies upon a mutual trust that sources, opinions, facts, and insights will be properly noted and carefully credited. In practical terms, this means that, as students, you must be responsible for the full citations of others' ideas in all of your research papers and projects; you must be scrupulously honest when taking your examinations; you must always submit your own work and not that of another student, scholar, or internet agent. Any breach of this intellectual responsibility is a breach of faith with the rest of our academic community. It undermines our shared intellectual culture, and it cannot be tolerated. Students failing to meet these responsibilities should anticipate being asked to leave Columbia.

Columbia College Honor Code:

The Columbia College Student Council, on behalf of the whole student body, has resolved that maintaining academic integrity is the preserve of all members of our intellectual community – including and especially students. As a consequence, all Columbia College students make the following pledge:

We, the undergraduate students of Columbia University, hereby pledge to value the integrity of our ideas and the ideas of others by honestly presenting our work, respecting authorship, and striving not simply for answers but for understanding in the pursuit of our common scholastic goals. In this way, we seek to build an academic community governed by our collective efforts, diligence, and Code of Honor.

In addition, all Columbia College students are committed to the following honor code:

I affirm that I will not plagiarize, use unauthorized materials, or give or receive illegitimate help on assignments, papers, or examinations. I will also uphold equity and honesty in the evaluation of my work and the work of others. I do so to sustain a community built around this Code of Honor.

If found guilty of cheating or plagiarism, you will receive a zero for that assignment and be sent to the Dean (www.college.columbia.edu/academics/disciplinaryprocess).) Please note that using code snippets from the internet IS acceptable, as long as you indicate where the code was copied from and provide a link to the public code source.

Citation should follow APA guidelines: http://www.apastyle.org/. If you have any doubt throughout the semester about how to cite something, or whether it would constitute as plagiarism, feel free to ask me.

Academic support services:

Writing Center - https://www.college.columbia.edu/core/uwp/writing-center

Columbia Libraries - http://library.columbia.edu/

The schedule and materials listed below are subject to minor changes based on the progression of the class.

Schedule and materials:

Week	Topics (Each class on the day a lab assignment is due will spend 30-45 minutes at the end of class reviewing the lab assignment)	Readings (To be read before the start of class each week, and discussed in class during the week they are listed for). (*Indicates reading is methodological and therefore will not be discussed in the same way as the research paper readings, rather it will be lectured on with questions taken.)	Assignments (due the following week)
1 09/14	-Course overview -Introduction to Behavioral Data Science	David Donoho (2017) 50 Years of Data Science, Journal of Computational and Graphical Statistics, 26:4, p745-766. Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., … & Alstyne, M.V. (2009). Computational social science. *Science*, 323(5915), p721-723. Adjerid, I., & Kelley, K. (2018). Big data in psychology: A framework for research advancement. American Psychologist, 73(7), p899-917. (Reading page count: ~42) *Optional*: Lazer, D. & and Radford, J. (2017). Data ex Machina: Introduction to Big Data. *Annual Review of Sociology*, p19-39. Note: this paper discusses Big Data and Sociology but in most cases "Sociology" can be replaced with "Behavioral Science".	Assn. 1: Review of R programming and statistical analysis Thought paper
DATA ACQUISITION			
2 09/21	-Big data and psychology -API data ingestion	Ruths, D., & Pfeffer, J. (2014). Social media for large studies of behavior. *Science*, *346*(6213), p1063-1064.* Murphy, S. C. (2017). A hands-on guide to conducting psychological research on Twitter. Social Psychological and Personality Science, 8(4), 396-412. (Please read the PDF posted on courseworks – it has some annotations that update some information in the paper).* Eichstaedt, J. C., Schwartz, H. A., Kern, M. L., Park, G., Labarthe, D. R., Merchant, R. M., ... & Weeg, C. (2015). Psychological language on Twitter predicts county-level heart disease mortality. Psychological science, 26(2), 159-169.	Assn. 2: API data ingestion Thought paper

		(Reading page count: ~38) *Optional*: Sisco, M.R., Bosetti, V., Weber, E.U. When do extreme weather events generate attention to climate change? Climatic Change 143 (1-2), 227-241	
3 09/28	-Social media data -Online search data -Big experiments	Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., & Watts, D. J. (2010). Predicting consumer behavior with Web search. *Proceedings of the National academy of sciences*, *107*(41), p17486-17490. Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, *111*(24), p8788-8790. (Reading page count: ~29) *Optional*: Sisco, M.R., Pianta, S., Weber, E.U., & Bosetti, V. (2020). Global climate strikes sharply raise attention to climate change: Analysis of climate search behavior in 46 countries. Under review at Journal of Environmental Psychology. *Will be uploaded to courseworks. Resilient cooperators stabilize long-run cooperation in the finitely repeated Prisoner's Dilemma A Mao, L Dworkin, S Suri, DJ Watts - Nature communications, 2017, p 1-9. [An example of "big experiments" only possible with online participants] Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. F., ... & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, *115*(37), 9216-9221. [A "small data" experimental study of social media effects] Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415), 295-298. [Another massive social media experiment, but here focusing on political behavior] Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. Economic record, 88, 2-9.	Assn. 3: Webscraping Thought paper
4 10/5	Data acquisition wrap up: -Google Trends data -Webscraping -Open databases	Landers, R. N., Brusso, R. C., Cavanaugh, K. J., & Collmus, A. B. (2016). A primer on theory-driven web scraping: Automatic extraction of big data from the Internet for use in psychological research. Psychological methods, 21(4), 475. * Wilson, G., Aruliah, D. A., Brown, C. T., Hong, N. P. C., Davis, M., Guy, R. T., ... & Waugh, B. (2014). Best practices for scientific computing. *PLoS biology*, *12*(1), e1001745, p1-6.* (Reading page count: ~43) *Optional:* Sisco, M. R., & Weber, E. U. (2019). Examining charitable	Assn. 4: R in the cloud Thought paper (this week you can also make it about ideas for your proposal since all papers are methodological)

		giving in real-world online donations. Nature communications, 10(1), 1-8.	
DATA PROCESSING			
5 10/12	Data processing: -Data processing overview -Parallel computing in R -Cluster computing -R in the cloud I (linux, FTP, SSH)	Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, *457*(7232), p1012-1014. Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: traps in big data analysis. *Science*, *343*(6176), p1203-1205. (Reading page count: ~7) Optional: Santillana, M., Nguyen, A. T., Dredze, M., Paul, M. J., Nsoesie, E. O., & Brownstein, J. S. (2015). Combining search, social media, and traditional data sources to improve influenza surveillance. PLoS Comput Biol, 11(10), e1004513. Rossini, A. J., Tierney, L., & Li, N. (2007). Simple parallel statistical computing in R. Journal of computational and Graphical Statistics, 16(2), 399-420. * Lane, R. (2019). Habanero – Getting Started and R Job Examples. HPC Cluster User Documentation. * (2 pages) Kane, M. J., Emerson, J., & Weston, S. (2013). Scalable strategies for computing with massive data. Journal of Statistical Software, 55(14), 1-19. * This paper presents more packages than we'll learn in this class, but they could be useful to know about. Strimas-Mackey, M. (2016). RStudio in the Cloud I: Amazon Web Services.* (2 pages)	Assn. 5: Parallel computing Thought paper
6 10/19	-Statistical programming for long-run analyses (batch programming, scheduling) -Estimating additional variables (gender, ideology, age) -SQL -Monte Carlo simulations (for statistical evaluations and power analyses)	Barberá, P. (2014). Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Political Analysis*, 23(1), 76-91. G Park, HA Schwartz, JC Eichstaedt, ML Kern, M Kosinski, DJ Stillwell, LH Ungar, ME Seligman (2015). Automatic personality assessment through social media language. Journal of personality and social psychology, 108(6). (Reading page count: ~28) *Optional:* Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. Proceedings of the National Academy of Sciences, 110(15), 5802-5805. Coppersmith, G., Dredze, M., & Harman, C. (2014, June). Quantifying mental health signals in Twitter. In Proceedings of the workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality (pp. 51-60). Matz, S. C., Schwartz, H. A., Menges, J. I & Stillwell, D. J.	Assn. 6: Estimating addnl. variables Thought paper

		(2019). Predicting Individual-level Income from Facebook Profiles. PLOS ONE, p1-13 Chapter 1 from Beaulieu, A. (2009). Learning SQL: Master SQL Fundamentals. O'Reilly Media, Inc. p1-14* Ferron, J. M., Farmer, J. L., & Owens, C. M. (2010). Estimating individual treatment effects from multiple-baseline data: A Monte Carlo study of multilevel-modeling approaches. *Behavior Research Methods*, *42*(4), 930-943.	
DATA ANALYSIS			
7 10/26	-Data analysis overview -Natural experiments (pre-post event analyses, IV regression, regression discontinuity)	Aral, S., & Nicolaides, C. (2017). Exercise contagion in a global social network. *Nature communications*, *8*, 14753, p1-8. (Implements an IV analysis). Kearney, M. S., & Levine, P. B. (2015). Media influences on social outcomes: The impact of MTV's 16 and pregnant on teen childbearing. American Economic Review, 105(12), 3597-3632. (Implements an IV analysis) (Reading page count: 37) *Optional*: Kearney, M. S., & Levine, P. B. (2019). Early childhood education by television: Lessons from Sesame Street. American Economic Journal: Applied Economics, 11(1), 318-50. (Natural experiment) Cappelleri, J. C., & Trochim, W. M. (2015). Regression discontinuity design. International encyclopedia of the social and behavioral sciences, 20. (This is written assuming you are planning a RDD but it also applies to naturally occurring RDDs. Sharma, A., Hofman, J. M., & Watts, D. J. (2015, June). Estimating the causal impact of recommendation systems from observational data. In Proceedings of the Sixteenth ACM Conference on Economics and Computation (pp. 453-470). ACM. (Implements an IV analysis). Sisco, M.R. & Weber, E.U. (2020). Local warming increases climate policy support: Analysis of survey responses, internet search frequencies, and US congressional vote shares. Paper under review at PNAS.	*No assn. or thought paper this week, instead prep for student proposals and presentations due next class*
8 11/9 (11/2 uni closed)	-Time series analyses (autocorrelation, stationarity, seasonality, unit roots, ARIMA) -Spatial analyses (spatial autocorrelation, spatial analysis programming, spatial regression modeling) -Student project proposals I	Chapter 1 of Pickup, M. (2014). Introduction to time series analysis (Vol. 174). Sage Publications.* (19 pages) Chapter 1 of Ward, M. D., & Gleditsch, K. S. (2008). Spatial regression models. Sage.* (33 pages) (Reading page count: 82) *Optional:* Chapter 2 of Pickup, M. (2014). Introduction to time series analysis (Vol. 174). Sage Publications.* (32 pages) Chapter 2 of Ward, M. D., & Gleditsch, K. S. (2008). Spatial regression models. Sage.* Chapter 3 of Ward, M. D., & Gleditsch, K. S. (2008). Spatial	Assn. 7: Time and spatial analyses Thought paper

		regression models. Sage.* Pianta, S. & Sisco, M.R. (2020). A hot topic in hot times: how media coverage of climate change is affected by temperature abnormalities. Environmental Research Letters. (10 pages)	
9 11/16	-Public opinion estimation (with surveys, digital data, and multiple regression with post stratification (MRP)) -Bootstrapping -Student project proposals II	*Pick at least one:* Wang, W., Rothschild, D., Goel, S., & Gelman, A. (2015). Forecasting elections with non-representative polls. *International Journal of Forecasting*, 31(3), 980-991 Beauchamp, N. (2017). Predicting and Interpolating State-Level Polls Using Twitter Textual Data. American Journal of Political Science, 61(2), 490-503. *Optional*: Klašnja, M., Barberá, P., Beauchamp, N., Nagler, J., & Tucker, J. (2017). Measuring public opinion with social media data. In *The Oxford handbook of polling and survey methods*. (33 pages) Kastellec, Jonathan P., Jeffrey R. Lax, and Justin Phillips. Working Papers. "Estimating State Public Opinion with Multi-level Regression and Poststratification using R" [Tutorial on how to implement MRP] Howe, P. D., Mildenberger, M., Marlon, J. R., & Leiserowitz, A. (2015). Geographic variation in opinions on climate change at state and local scales in the USA. Nature Climate Change, 5(6), 596-603. [Another great example using MRP]. (Reading page count: 58)	Assn. 8: Public opinion estimation Thought paper
10 11/23	-Machine learning (core concepts, popular classical and modern models, performance evaluation) **Guest presenter: Tal Golan (neural networks models)*	Chapters 1, and 2.1-2.3 of James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning. New York: Springer, p1-37.* Gladstone, J. J.*, & Matz, S. C.* (2019). Can Psychological Traits be Inferred from Spending? Evidence from Transaction Data. Psychological Science, p1087-1096. Bhatia, S. (2019). Predicting risk perception: new insights from data science. Management Science. (23 pages) Blumenstock, J., Cadamuro, G., & On, R. (2015). Predicting poverty and wealth from mobile phone metadata . *Science*, *350*(6264), 1073-1076. (Reading page count: 75) *Optional*: LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.	No lab assn. this week, use the time to work on your final project. Thought paper
11 11/30	-Regression in machine learning (automated model selection, LASSO regression)	Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267-288.* Reece, A. G., Reagan, A. J., Lix, K. L., Dodds, P. S., Danforth, C. M., & Langer, E. J. (2017). Forecasting the onset and course of mental illness with Twitter data. Scientific reports, 7(1), 13006. (9 pages) Bedi, G., Carrillo, F., Cecchi, G. A., Slezak, D. F., Sigman,	Assn. 9: Basic ML and ML with regression Thought paper

		M., Mota, N. B., ... & Corcoran, C. M. (2015). Automated analysis of free speech predicts psychosis onset in high-risk youths. *npj Schizophrenia*, *1*, 15030. (6 pages) Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. Perspectives on Psychological Science, 12(6), 1100-1122. (Reading page count: 59)	
12 12/7	-Natural Language Processing I (word counting, feature extraction, NLP and machine learning, ensemble models) **Guest presenter: Sarah Ita Levitan (Linguistic analysis of mental illness)*	Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology*, 29(1), 24-54. Doré, B., Ort, L., Braverman, O., & Ochsner, K. N. (2015). Sadness shifts to anxiety over time and distance from the national tragedy in Newtown, Connecticut. Psychological science, 26(4), 363-373. Coppersmith, G., Dredze, M., Harman, C., & Hollingshead, K. (2015). From ADHD to SAD: Analyzing the language of mental health on Twitter through self-reported diagnoses. In Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality (pp. 1-10). Yancheva, M., Fraser, K., & Rudzicz, F. (2015). Using linguistic features longitudinally to predict clinical scores for Alzheimer's disease and related dementias. In Proceedings of SLPAT 2015: 6th Workshop on Speech and Language Processing for Assistive Technologies (pp. 134-139). Linguistic Analysis of Schizophrenia in Reddit Posts Jonathan Zomick, Sarah Ita Levitan, and Mark Serper CLPsych 2019 (Reading page count: 61)	Assn. 10: Basic natural language processing Thought paper
13 12/14	-Natural Language Processing II (topic modeling, word vectors) -Student project presentations	Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84. Skim: Spirling, A. and Rodriguez, P. (2019). Word Embeddings What works, what doesn't, and how to tell the difference for applied research. Working paper, New York University (29 pages) Bhatia, S. (2017). Associative judgment and vector space semantics. Psychological Review, 124(1), 1. (18 pages) Bhatia, S. (2016). Vector Space Semantic Models Predict Subjective Probability Judgments for Real-World Events. In CogSci. (6 pages) (Reading page count: 56)	Final paper due December 21st.