STA 4930 – Data Science Capstone Project

Course Description: This culminating experience will synthesize methods and ideas acquired in undergraduate courses by working in small groups on a project and presenting findings in a written report, poster presentation, and an oral presentation.

Required Prerequisites: Permission of the instructor

Instructor: Dr. Elizabeth Johnson   ***Email****:* ejohnson5@ufl.edu
Office: 115-B Griffin-Floyd Hall   ***Telephone***: 352-273-1897

Class Meeting Times: Thursdays from 1:55-3:50 (classroom TBD) and one additional virtual period TBA.

Course Goals: The primary goal for this course is to provide a unifying experience that links statistical concepts across courses and provides a framework for conducting independent research. We will further develop the skills in exploring data, building, and fitting models, investigating model assumptions, interpreting results from statistical models, and report writing.

Learning Outcomes: After completion of this course students will be able to:

- integrate their knowledge from previous courses to conduct an original research project

- build collaborative skills by working with professionals in related fields

- develop effective communication (written and oral) skills by presenting the results of their research to both technical and non-technical audiences.

- demonstrate competence in using statistical software.

Selected Readings and Resources*:*

Selected course readings related to the study, such as journal articles, will be posted on the Canvas site.

Hadley Wickham and Garrett Grolemund , 2017. R for Data Science, O'Reilly, Addison Wesley Download free pdf at [Welcome | R for Data Science (had.co.nz)](had.co.nz)

Gareth James, et al. 2013. An Introduction to Statistical Learning with Applications in R, Springer. Free pdf download available at [An Introduction to Statistical Learning (statlearning.com)](statlearning.com) Access to slides and 15 hours of lecture videos available at [In-depth introduction to machine learning in 15 hours of expert videos | R-bloggers](R-bloggers)

Software: R and selected R packages constitute the primary software for this class. The Comprehensive R Archive (CRAN) is the primary place to download R. RStudio is the recommend integrated development environment for using R. RStudio Download. The Free Desktop version is fine. Other software packages may also be utilized.

Student Evaluation: Student work in the course (and the relative weighting of this work in the overall grade) will consist of:

Class readings and discussions	15%
Background research	15%
Research Paper	50%
Poster	15%
Oral Presentation	10%

More information about each of these assignments is listed later in the syllabus. The instructor reserves the right to adjust the percentages if needed.

Grading Scale:

Numeric Score	Letter Grade	Numeric Score	Letter Grade
93 – 100	A	77– 79	C+
90 – 92	A–	71 – 76	C
87 – 89	B+	67 – 70	C–
83 – 86	B	60 – 66	D
80 – 82	B–	0 – 59	E

Course Assignments: Your final course grade will be based on a variety of assignment types that are described in detail below. Due dates will be posted on the course schedule in Canvas and announced in class.

Class readings and discussions: Participation is an important component of this course. You are expected to participate in all in-class exercises and discussions by completing the required readings or activities.

Background Research: This may involve activities such as: attending professional presentations, either in-person or online, finding resources in the library, or meeting with other professionals in the field of the study. Students will write a one-page summary of each chosen activity. Summaries of professional talks should include the following items: speaker information – title and affiliation, title of talk, overall research question(s), type and method of data collected, type of statistical methodology used or discussed, overall presentation style.

Capstone Research Project: Student group(s) will complete their research project on their approved topic. Based upon your work, your group will write a statistical paper, usually approximately 10-15 pages, prepare a poster, and give an oral presentation to an audience of faculty and peers. See rubrics below.

Rubric for STA 4930 Oral Presentation

	1	2	3	4
Organization	Talk lacks any semblance of logical structure.	Relationship between portions of talk is unclear AND student understanding is compromised by this lack.	Occasionally the relationship between parts of the talk is unclear OR the organization of the talk, while existent, is not conducive to student understanding.	Logical development of talk is clear and appropriate. Organization of talk enables students to understand the material.
Clarity of communication	The talk is confusing throughout.	The theme of the talk is unclear OR many points in the talk are unclear. Incorrect statistical language is used.	Speaker generally communicates clearly, but some fine points are unclear, or some incorrect statistical language is used.	Speaker communicates coherently and clearly, and precise statistical language is used.
Representations (graphs, tables, equations, …) and notation	Speaker uses unclear or inappropriate representations, or speaker uses no representations whatsoever.	Speaker uses minimal representations, or uses representations ineffectively.	Speaker effectively uses clear and appropriate representations, including technology, manipulatives, handouts, and/or writing on board.	Speaker effectively uses clear and appropriate representations, including technology, manipulatives, handouts, and/or writing on board, and addresses connections between different representations.
Knowledge of subject matter	Speaker shows errors in knowledge of statistical concepts.	Speaker explains concepts but in a rudimentary form.	Speaker clearly articulates statistical concepts.	Speaker clearly and coherently articulates statistical concepts and develops connections among concepts.
Mathematical proof or statistical methodology	Speaker does not understand proof or methodology	Speaker lacks appropriate reasoning and methods in explaining proof or methodology	Speaker explains proof/methodology with acceptable reasoning and methods, and partially supplies detail.	Speaker thoroughly explains proof/methodology using sound reasoning, appropriate methods, and supplies pertinent detail.

Grading Expectations: Written Report Checklist

Your report will be graded according to how you meet the criteria detailed below:

Report Goal
Do you have a clearly defined goal or goals?
Intended Readers
What are the backgrounds of the intended readers?
What are they likely to know about the topic?

Title
Does the title completely describe the subject of the report?
Are there any unnecessary words?

Abstract
Is the abstract understandable without reading the rest of the report?
Does it summarize the objectives, results and implications of your study?

Introduction
Can the reader understand why the research was done?
Can the reader understand how the research fits into existing knowledge of the upper-level STA course topic?
Does the final paragraph outline the remainder of the report?

Methods and Results
Are the statistical methods and results clearly presented?
Are all tables and figures numbered, titled and clearly introduced in the prose?
Do the tables and figures tell their part of the story?

Discussion
Is your discussion consistent with the statistical results?
Have you clearly distinguished conjecture from fact?
Is your discussion free from statistical jargon?
Are the implications of your study clearly discussed?
Have you pointed out any shortcomings or limitations in your study?

Conclusion
Have you summarized the main points of your report?
References/Appendices
Have you given complete citations for all sources referenced in your report?
Are the citations consistent with the style you are using?
Are the appendices appropriately titled, identified by a letter or number and referred to in the body of the report?

Final Checks
Are the pages numbered?
Have you used a spell checker?
Have you carefully proofread your report after making all the changes?
Has a colleague critiqued your report?

Poster Presentation Evaluation Criteria

Presenter(s): _____ **Reviewer:** _____

Please rate the poster/presenter utilizing the 0 to 3 scale where
0=No attempt, 1=Developing, 2=Competent and 3=Exemplary

1. Statement of Research Problem/Rationale:

a) Clearly written questions or hypotheses being addressed	0	1	2	3
b) Well-written rationale/justification for the study	0	1	2	3
c) Information includes an explanation of key concepts/theories	0	1	2	3

2. Methods (Explanation/Appropriateness):

a) Clear description of data collection procedures if applicable	0	1	2	3
b) Clear description of statistical methods used	0	1	2	3
b) Methods are appropriate to address aim/question	0	1	2	3
c) Sufficient written information to suggest that statistics are calculated correctly	0	1	2	3

3. Analysis/Results:

a) Figures/Tables used to appropriately and clearly present the data	0	1	2	3
b) Findings are presented clearly and accurately	0	1	2	3
c) Analysis is well written and appropriately applied	0	1	2	3

4. Written Conclusion/Discussion:

a) Addressed study's problem/question	0	1	2	3
b) Conclusions are sufficiently supported by results	0	1	2	3
c) Results are placed into broader framework	0	1	2	3
d) Importance of findings is addressed	0	1	2	3

5. Verbal Interaction with Presenter(s)

a) Presenter(s) summarized study clearly	0	1	2	3
b) Length of poster summary by presenter(s) was appropriate	0	1	2	3
c) Presenter(s) demonstrated full knowledge of the material and can explain and elaborate on questions	0	1	2	3

6. Poster Design

a) Effective overall aesthetic /organization of poster	0	1	2	3
b) Flow of information is logical and facilitates understanding	0	1	2	3
c) Fonts are easy to read and vary appropriately for headings and text	0	1	2	3
d) Correct grammar and spelling	0	1	2	3

7. Overall

a) Question is interesting	0	1	2	3
b) Study is original	0	1	2	3
c) Potential to contribute to the field	0	1	2	3

STA 4930 Research Paper Rubric

Name: _____

Team Members: _____

Title: _____

	1- Minimal	2- Developing	3 - Substantial	Score
1. Does the title give an accurate preview of what the paper is about?	Poor and uninformative title	Imprecise title – too generic	Title is informative, specific, and precise	
2. Does the abstract summarize the paper?	The abstract does not provide an effective summary.	The abstract does not mention some key elements in the paper such as results	The main points of the paper are described clearly and succinctly.	
3. Does the background and significance of the study have a logical organization? Does it move from the general to the specific?	Poor Progression - Difficult to follow train of thought - Results presented out of order Poor Lexicon - Uses lots of unclear jargon, phrases or terms that would be unfamiliar to the audience	Fair Progression - Fairly clear train of thought - Results presented in reasonable order Fair Lexicon - Uses little unclear jargon, phrases or terms that would be unfamiliar to the audience	Great Progression - Intuitive order to presentation materials - Covered all important topics Great Lexicon - Uses no jargon that would be unfamiliar to audience - Clear/concise language used.	
4. Has sufficient background been provided to understand the paper?	No background or irrelevant background provided	Minimal background but lacking in specifics.	This work clearly relates to other work in the scientific literature.	

	1- Minimal	2- Developing	3 - Substantial	Score
5. Has a reasonable explanation been given for why the research was done? Why is the work important? Why is it relevant?	The paper does not contain statements about the hypothesis/goals of the paper Poor Strength of Story - Did not convince audience of importance/impact	The paper contains unclear statements about the hypothesis/goals of the paper Fair Strength of Story - Convinces audience of importance/impact, but could be stronger	The paper contains clear statements about the hypothesis/goals of the paper Exceptional Strength of Story- convinces audience of importance/impact	
6. Is the data collected or obtained in an appropriate manner for this study?	Data were not collected/obtained in an appropriate manner to answer the statistical question	Data were collected/obtained in an appropriate manner to answer the statistical question, but raw data are not included	Data were collected/obtained in an appropriate manner to answer the statistical question. Raw data are included	
7. Is the selection and creation of variables clearly described?	Variables and experimental units not clearly defined.	Variables and experimental units are defined but the study could not be repeated based on the information provided	Clear indication of the experimental unit and both the response and explanatory variables including variable type The study could be repeated based on the information provided	
8. Is the analytic method used appropriate for the study?	Inappropriate statistical method selected, or an appropriate method incorrectly described.	Appropriate statistical method selected but not clearly described.	Appropriate statistical method selected and correctly described.	
9. Are the data visualizations (including tables and graphs) impactful?	Poor Depth - Only very simple data attributes plotted Poor Impact - Data visualization did not contribute in a meaningful way to the presentation	Fair Depth - Relatively simplistic data relationships visualized Fair Impact - Data visualizations contribute moderately to the presentation	Includes appropriate, well-labeled, accurate displays (graphs and tables) of the data Great Impact - Data visualizations contribute in a major way to the presentation	

	1- Minimal	2- Developing	3 - Substantial	Score
10. Is the analysis of the data appropriate and complete?	Analysis of the data is not appropriate or is missing	Analysis of the data is appropriate, somewhat accurate, and fairly thorough	Analysis of the data is accurate, thorough, and appropriate Conditions are checked correctly for any inference procedures	
11. Do the results address the research question (support or disprove a hypothesis)?	The author does not address the research questions.	The author articulates the basis for supporting or rejecting each hypothesis but does not use specific data cited from the results.	The author clearly articulates the basis for supporting or rejecting each hypothesis using specific data cited from the results.	
12. Does the author adequately discuss any limitations of the research and discuss possible future studies?	No evidence of reflective thinking	Reflective thinking focuses on the positive aspects of the project only	Gives a good overall picture of the project—what went well and what didn't—and includes ideas for further study	
13. What is the overall impression of the paper?	Project report has none of the three qualities: attractive, well-organized, well-written. References are either not included or are improperly cited	Project report has one of the three qualities: attractive, well-organized, well-written Appropriate references of adequate quality are cited properly (both in the text and at the end of the paper)	Attractive, well-organized, well-written project report. Appropriate references of adequate quality are cited properly (both in the text and at the end of the paper)	

Notes:

1) *A score of zero will be assigned to any of the above sections that do not meet the criteria of a score of 1- minimal.*
2) *Individual team members will be scored both on the overall paper and on their own specific contribution to the analysis of the data.*

Course Policies

Grading Policies:

*Requirements for class attendance and make-up exams, assignments, and other work in this course as well as policies regarding absences, religious holidays, illness, and student athletes are consistent with UF Attendance Policies.

Additional make-up policy requirements:
- Every effort should be made to complete the assignment/exam during the open period. Only extreme situations will warrant a makeup. Contact the instructor prior to the exam - as soon as you realize you will be unable to take the assignment/exam at the scheduled time. Each case will be reviewed individually. Valid and detailed documentation is a prerequisite for scheduling a makeup under such extenuating circumstances.
- If you have an emergency on the day of the assignment/exam, the instructor must be contacted by midnight of the day of the assignment/exam.
- Make-up exams will be scheduled within a week from the assignment deadline. Student is responsible for attending scheduled make-up. Instructor reserves the right to utilize the UF posted final exam day as a make-up date.
- Additional Note: Being on vacation or booking a trip prior to the completion of the semester is not a valid reason to request a makeup. Please reference the most recent Academic Calendar, https://catalog.ufl.edu/UGRD/dates-deadlines/pdfs/.

*If you have a disability that requires academic accommodation, contact the Disability Resource Center (DRC). The DRC will provide documentation to the students who must then provide this documentation to the instructor when requesting information. You must submit this documentation prior to submitting any assignments for which you are requesting accommodation.

* Incompletes are only assigned when extraordinary circumstances (such as an accident, or extended hospitalization), arising after the date for dropping the course, prevent the student from completing the course requirements. Having a failing grade in the course is not a valid reason for requesting an Incomplete. Information on Medical Withdrawal can be found at https://umatter.ufl.edu/ . Information on how to Drop a class can be found in UF's Academic Catalog https://catalog.ufl.edu/ and https://catalog.ufl.edu/UGRD/academic-regulations/dropping-courses-withdrawals/

Classroom Behavior: During class students should silence their cellular phones and refrain from eating, drinking, reading newspapers, doing homework, listening to music, excessive talking and all other behaviors that are distracting and disrespectful to the instructor and their fellow students.

Privacy Policy: Student records are confidential. Only information designated "UF directory information" may be released without your written consent. This applies to parents or anyone else who contacts me about your grades.

Faculty Course Evaluations: Student feedback is welcomed by the instructor and beneficial to future students in the course. Students are requested to provide feedback on the quality of instruction in this course by completing a brief confidential evaluation towards the end of the semester at https://evaluations.ufl.edu. Summaries of the evaluation results can be found at https://evaluations.ufl.edu/results.

Other University Services

U Matter, We Care: Your well-being is important to the University of Florida. The U Matter, We Care initiative is committed to creating a culture of care on our campus by encouraging members of our community to look out for one another and to reach out for help if a member of our community is in need. If you or a friend is in distress, please contact umatter@ufl.edu so that the U Matter, We Care Team can reach out to the student in distress. A nighttime and weekend crisis counselor is available by phone at 352-392-1575. The U Matter, We Care Team can help connect students to the many other helping resources available including, but not limited
to, Victim Advocates, Housing staff, and the Counseling and Wellness Center. Please remember that asking for help is a sign of strength. In case of emergency, call 9-1-1.

*Sexual Assault Recovery Services (SARS): Student Health Center, 392-1161
*University Police Department, 392-1111 (or 9-1-1 for emergencies), http://www.police.ufl.edu

*Student Health Care Center: Call 352-392-1161 for 24/7 information to help you find the care you need or visit the Student Health Care Center website.

*GatorWell Health Promotion Services: For prevention services focused on optimal wellbeing, including Wellness Coaching for Academic Success, visit the GatorWell website or call 352-273-4450.

Academic Resources

E-learning technical support: Contact the UF Computing Help Desk at 352-392-4357 or via e-mail at helpdesk@ufl.edu.

Writing Studio: 2215 Turlington Hall, 352-846-1138. Help brainstorming, formatting, and writing papers.

Student Complaints On-Campus: Visit the Student Honor Code and Student Conduct Code webpage for more information.

COVID -19

In response to COVID-19, the following recommendations are in place to maintain your learning environment, to enhance the safety of our in-classroom interactions, and to further the health and safety of ourselves, our neighbors, and our loved ones.

- If you are not vaccinated, get vaccinated. Vaccines are readily available and have been demonstrated to be safe and effective against the COVID-19 virus.

- If you are sick, stay home. Please call your primary care provider if you are ill and need immediate care or the UF Student Health Care Center at 352-392-1161 to be evaluated.

- As with any excused absence, you will be given a reasonable amount of time to make up missed work.