MERRIMACK COLLEGE
SCHOOL OF ENGINEERING & COMPUTATIONAL SCIENCES

Visual Data Exploration: DSE5004

Instructor: Dr. Michael Dupin
Instructor Email: dupinm@merrimack.edu
Course Term: Spring 2 2024
Synchronous Live Session: Tuesdays, 7 to 8:30 PM ET
Office Hours: Monday - Friday, by appointment

Course Description

This course introduces students to the fundamentals of exploratory data analyses and data visualization, allowing students to turn data into insights, pictures, and stories. The following topics will be covered:

- Transforming data into information and subsequent actionable insights and knowledge
- Exploring the basics of graphic design and making a "good" graph
- Undertaking data due diligence
- Descriptive and inferential data analytical framework
- Examples and practice of exploratory data analysis and using data for segmentation and classification
- Exploring why some data visualizations present information effectively and others do not,
- Considering visualization as a component of the data analytics systems

The course will address data exploration utilizing classical statistical techniques, as well as topics of statistical inference, statistical significance, and outcome validity and reliability. Additionally, this course will introduce students to the foundations of constructing plots with the grammar of graphics and utilizing packages such as ggplot2 and Shiny.

Course Learning Objectives

Upon completion of this course, students should:

- Distinguish between data, information, and knowledge.
- Understand the basic data types and effective visual mappings
- Relate key design guidelines and techniques used for the visual display of information, including their relationship to human perception.
- Exhibit sound command of the data due diligence process, data exploration, and interpretation of results.
- Apply sound statistical reasoning to create evidence for a given scenario.
- Conduct basic data due diligence and exploration activities in a case study.
- Apply fundamental data visualization techniques and theoretical principles to real-world problems from exploratory data analysis to dynamic presentations for decision making.
- Design static and interactive visualizations to effectively present information to users and/or to enable data exploration, using real data and a human-centered process.

Important Deadlines for the Term

03/11/2024	Monday	First day of classes
03/15/2024	Friday	Last day to add/drop a course
04/19/2024	Friday	Last day to withdraw from a course with a "W"
05/03/2024	Friday	Last day of classes
05/07/2024	Tuesday	Grades Due

Required Course Materials

Yau, Nathan. 2011. *Visualize This: The Flowing Data Guide to Design, Visualization, and Statistics.* John Wiley & Sons, Inc. , ISBN-13: 978-0-470-94488-2

Cairo, Alberto. 2013. *The Functional Art: An Introduction to Information Graphics and Visualization*. New Riders, Pearson. ISBN-13: 978-0321834737

Optional Course Materials (will be available via PDF or book website)

Books

Banasiewicz, Andrew. *Marketing Database Analytics: Transforming Data for Competitive Advantage*, Erudite Systems, ISBN-13: 978-0415657884

Provost, Foster and Tom Fawcett. 2013. *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*, O'Reilly, ISBN-13: 978-1449361327

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*, Springer, 2nd Edition, ISBN-13: 978-3319242750. Work in Progress Online Link: https://ggplot2-book.org/

Articles

Hagiu, Andrei and Julian Wright. 2020. When Data Creates Competitive Advantage. Harvard Business Review, https://hbr.org/2020/01/when-data-creates-competitive-advantage

Gallo, Amy. 2014. A Refresher on Statistical Significance. Harvard Business Review, https://hbr.org/2016/02/a-refresher-on-statistical-significance.

Svolba, Gerhard. 2019. 7 Ways Analytical Methods Improve Data Quality. SAS. https://blogs.sas.com/content/hiddeninsights/2019/04/02/7-ways-analytical-methods-improve-data-quality/

Tableau Download

For the final project, you will be required to use Tableau. You can request a student license on the Tableau website: https://www.tableau.com/academic/students using your Merrimack license.

Grading & Assignments

A Excellent	B Good	C Fair/Poor	F Unacceptable
93+ A	87-89 B+	77-79 C+	Below 70
90-92 A-	83-86 B	73-76 C	
	80-82 B-	70-72 C-	

Although the graduate grading policy is similar to the undergraduate policy, it should be noted that the expectations for graduate students are much higher and therefore the grading is more rigorous. Candidates for any graduate-level degree or certificate must attain a final cumulative grade point average of 3.0 before the degree or certificate will be conferred.

The following letter and special grades are used across all graduate programs. Please note that this document sets the minimum standards for the College. Individual degree programs may have stricter GPA and course grade requirements.

A: "A" indicates outstanding work

B: "B" means that the work is satisfactory

C: "C" (2.0) is deemed unsatisfactory at the graduate level. No more than two courses at the C level (2.0 or higher) will be counted as acceptable toward a graduate degree. Students may be permitted to repeat only two courses and may repeat each course only one time. Those who receive more than two C level grades will be automatically dismissed from their program of study.

Any grade lower than a C - will not be acceptable for graduate-level work and cannot be counted as credit towards the degree. However, the grade will be counted toward the graduate student's GPA. If a student receives a grade lower than a C in a required course, the student will be required to retake the course. A course may only be retaken one time. The most recent grade will replace the first grade and will be factored into the student's GPA.

Please see "Academic Requirements and Policies" in the Graduate Catalog at https://catalog.merrimack.edu/index.php for more information.

Student requirements and their corresponding weights are summarized as follows:

Assignment	Weight
Introduction	5%
Quizzes	10% (1.25% each)
Position Papers (2)	20% (10% each)
Project 1: Data Due Diligence	25%
Project 2: Segmentation and Profiling	40%

All assignments will be graded on a standard 100-point scale, with each weighted in accordance with the schedule depicted above. Please note that there are two (2) Position Papers, thus each will be worth 10% of the final grade (for a total of 20%). There will be a small quiz on materials in each module. The quizzes are designed to track progress and assess items that might need to be reinforced at a later date. Please note that ALL of the following assignments are INDIVIDUAL – i.e., are to be completed individually by each student.

If you use a large language model tool (ChatGPT or other variants) for your research, I expect you to check all references provided by such tools before you include them in your writing. Incorrect or made-up references will result in a grade of 0 for the assignment.

Grading-Related Policies for this Course

All course assignments are to be submitted electronically to Canvas. There will be NO late work accepted. Final grades will be based on a percentage of possible points earned; but the instructor reserves the right to give a curve to adjust the average course grade.

Class Policies

1. Academic Integrity

See Merrimack College Student Handbook for college's Academic Integrity Policy. The academic integrity code and policy is also posted in the Graduate Catalog: https://catalog.merrimack.edu/content.php?catoid=9&navoid=202#academic-integrity
Here is a brief excerpt from that: "…Academic integrity is fundamental to creating and maintaining an atmosphere of cooperation and trust. It is thus a concern for everyone in the college community. The academic integrity code below is designed to help students understand what is not permissible in their academic and intellectual lives at the college. It seeks to protect students from unintentional acts of dishonesty and to preserve the trust inherent in the student- teacher relationship, which is compromised if suspicion arises regarding the integrity of a student's work. The code is also designed to inform students of the rules which will be used to judge academic integrity infractions…"

Specifically concerning the production of code it is important to recall that using code, text or other forms of media, such as homework solutions, from previous offerings of the course is forbidden, except as a "reference" or "guide" to work from, and even as such, **it is necessary to explicitly cite all consulted sources**. Tutors or other students who have taken the course in the past may help current students, but should not show them the solutions that they have developed for assignments. The lack of detailed citation of previous offerings of the course or any other similar sources, notably from the Internet or Generative AI systems, will be considered as submission of copied material, and a serious breach of academic integrity. Simply put, any work that is not your own should be clearly identified either through citations in written work or in comments in submitted files. This includes not only things that are directly copied from other's work, but also for ideas and procedures gathered from outside sources. It is not necessary to cite the materials presented in class.

2. Academic Accommodations from the Accessibility Services Office

Merrimack College provides reasonable accommodations for students with documented disabilities. Students who have, or think they may have, a disability are invited to contact the

Accessibility Services Office via the online request form found on the Accessibility Services website: www.merrimack.edu/aso, email: accessibilityservices@merrimack.edu, or by visiting us on the third floor of McQuade Library (in the event we are open).

Students are encouraged to contact the office as soon as possible to ensure adequate time to meet and create a plan. Students already registered with Accessibility Services are encouraged semesterly to request for their letters to be emailed and students are responsible to then email the PDF to their instructors personally. Accommodations cannot be made retroactively.

3. Live On-line Meeting Attendance

Participation in weekly live on-line meetings is highly encouraged, but it is not mandatory. All sessions will be recorded, and recordings will be posted; however, when not attending students forego the ability to directly ask questions about topics being discussed, or other parts of the course.

4. Communication

I will be communicating through official college email (merrimack.edu) and Canvas announcements as needed. Those will be the primary methods for dissemination of information between us, so please check your email at least once a day. I check my email several times a day, so it is the quickest way to reach me with questions. I will also be archiving important announcements through Canvas. When emailing me, use your Merrimack account, include a greeting with the class code and a closing signature with your name, or I will likely not respond.

If you are having trouble with an assignment, please do not suffer in silence. Let me know if you are struggling, either by email or by scheduling an appointment. I'm interested in seeing students succeed and am more than happy to provide extra help.

5. Requests for Extensions

The general policy is that, outside of properly verified serious medical emergencies* (as defined below), extensions are not given, which applies to all 4 assignments. Missing an assignment without an acceptable reason (to be clear, that means a serious medical emergency, as defined below) will result in 0 points for the exam or a project. The intent here is not to penalize anyone – quite to the contrary, it is to create a level playing field so that no one has a unique and an unfair advantage. All assignment due dates are published (see the Live Meetings, Topics & Assignments section below) and will not change, barring a natural or other emergency – please consider those dates when planning any non-class related activities.

**Serious medical emergency is defined as an injury or illness that is acute and poses an immediate risk to a person's life or long term health. To be "properly verified", the said serious medical condition must be attested to by hospitalization and related medical treatment documentation*.

6. Mental Health

It is important to make your mental health a priority! We will do activities in class to support

your mental health and wellness, and I also encourage you to engage in your own self-care habits outside of class. If you want more information or resources, please come see me! If you are struggling with your mental health, or you believe a classmate is, please talk to me so I can put you in touch with qualified and caring support to get you back on track and feeling better.

7. **Suicide Prevention Lifeline:**
   We can all help prevent suicide. The Lifeline provides 24/7, free and confidential support for people in distress, prevention and crisis resources for you or your loved ones, and best practices for professionals.
   http://www.ulifeline.org/stay_well
   National Suicide Prevention Lifeline 1-800-273-8255(TALK) or Call or Text 988

8. **Counseling Services:**
   Counseling Services are available to Graduate students at the Counseling Center at Merrimack via Uwill for telecounseling services only. Telecounseling includes video, chat, phone, message. Uwill is HIPPA and FERPA compliant which means they follow the strictest privacy guidelines. To access Uwill, register with your Merrimack email address at app.uwill.com. To receive support after hours call 978-837-5444 to be connected to the Uwill 24/7/365 crisis line which offers phone assistance only. The Counseling Center is located on the third floor of the Sakowich Center.

9. **Student Success Resources:**
   It takes a village to progress in your knowledge. The numerous resources available at Merrimack College to assist you in meeting your goals are outlined in the **Data Science Hub in Canvas**.

   For assistance in navigating the student success resources please reach out to the graduate advising team via ecs-grad-advising@merrimack.edu

Weekly Topics & Assignments

Week	Data Exploration	Data Visualization	Readings & Assignments
1	Introductions Course learning objectives Understand the distinction between data, information and knowledge Analytic maturity	Introduction to Data Visualization	Lecture slides PF Ch. 1 B Ch. 1 Y Ch. 1 C Ch. 1-3 **Quiz 01**
Introduction due			
2	Data as a source of competitive advantage CRISP framework Data Science Value Chain	Human Perception/ Cognition	Lecture slides PF Ch. 2 B Ch. 3 & 6 C Ch. 5-7 Y Ch. 5-7; W Ch. 1-2 Quiz 02
Position Paper 1 due			
3	Deep dive into the Data Science Value Chain Introduction to Machine Learning	Grammar of Graphics	Lecture slides PF Ch. 13 B Ch. 8 W Ch. 3-4 **Quiz 03**
Position Paper 2 due			
4	Plunging into data - sources, dimensions, types, typologies, and schemas *Overview Data Due Diligence project*	Grammar of Graphics (ggplot2)	Lecture slides PF Ch. 11 **Quiz 04** *Data Due Diligence project*
5	Exploratory Data Analysis (EDA) process Data diagnosis and treatments *Project progress check in and guidance*	Special Cases: Time, Space, Clusters	Lecture slides PF Ch 3 B Ch 5 & 7 W Ch 5 - 6 Y Ch 4 **Quiz 05** *Data Due Diligence project*
Data Due Diligence Project due			
6	Segmentation – common schema and methodologies *Overview of Segmentation and Profiling project*	R Shiny	Lecture slides PF CH. 5 - 6 Shiny tutorial **Quiz 06** *Segmentation project*
7	Segmentation methodologies *Project progress check-in and guidance*	Dashboard Creation	Lecture slides B Ch 13 - 14 W Ch 7 - 8 **Quiz 07** *Segmentation project*
8	Course wrap-up *Final progress check-in and guidance*	Visualization Tools	Lecture slides **Quiz 08** *Segmentation project*
Segmentation & Profiling Project due			