University of Wisconsin-Stevens Point

College of Professional Studies

School of Business and Economics

Spring 2020 - Version 1.10 (1/20/2020)

Course: Introduction to Data Analysis (DAC 101) #42385

Books: Introduction to Data Science (2017) by Jeffrey S. Saltz & Jeffrey M. Stanton.

(ISBN-13: 978-1506377537*) Available through the bookstore*

Book Website https://study.sagepub.com/saltz

Class Time: TR 12:00-1:15 p.m. (Room: SCI B228)

Professor: Dr. Kurt A. Pflughoeft (Floog’heft)

Office: CPS 442

Office hours: MT: 2-3 pm, W: 11-12 and by appointment

Contact: [kpflugho@uwsp.edu](mailto:kpflugho@uwsp.edu)

**Course Description:** Introduction to the concepts of how different areas make use of data analytics, and the stages of analyzing data; learn to interpret data, communicate insights, and leverage data. The course demonstrates applied examples of data collection, processing, transformation, management, and analysis to provide students with a hands-on introduction to the data analysis experience. Students will explore key concepts related to data analysis, including applied statistics, information visualization, text mining, and machine learning. “R”, the open source statistical analysis and visualization system, is used throughout the course. R is reckoned to be the most popular choice among data analysts worldwide; having knowledge and skill with using it is considered a valuable and marketable job skill for data scientists.

Week # Dates Topic\* Reading/Assignments

1. Jan 21, 23 Data Science Intro Ch 1-3, Lab 1, 2

*Video: Ch 3: Intro to R coding*

1. Jan 28, 30 Structured Data Ch 4-6, Lab 3,4 HW 1

*Videos: Ch 5: Rows & Columns, Ch 6: Sorting Dataframes*

1. Feb 4, 6 Summarization and functions Ch 7-9, Lab 5,6 Q1

*Videos: Ch 8 Functions, Ch 9 Distributions*

1. Feb 11, 13 Sampling and Inference Ch 10, Lab 7,8, HW 2

*Videos: Ch10 Samples, Sampling*

1. Feb 18, 20 External Data . Ch 11, Lab 9,10, Q2

*Videos: Ch11 SQL, JSON*

1. Feb 25, 27 Visualization Ch 12, Lab 11,12

*Videos: Ch 12 Basic GGPlot, Bar/Scatter Plots*

1. Mar 3, 5 Mapping Ch 13, Lab 13,14, Q3

*Videos: Ch 13 Points on a Map, Map Zooming*

1. Mar 10, 12# Review, #Midterm Exam .

Mar 17, 19 SPRING BREAK

1. Mar 24, 26! Linear Relations Ch 16, Lab 15,16 HW 4.

*Video: Linear Modeling, ! may be an assigned lab*

1. Mar 31, Apr 2 Rules Ch 17, Lab 17,18, Q4

*Video: Rule Mining*

1. Apr 7, 9 Vectors Ch 18, Lab 19,20 HW 5.

*Video: Ch19 SVM*

Week # Dates Topic\* Reading/Assignments

1. Apr 14, 16 Predictions, Text mining Q5, Lab 20,21

*Video: Ch 14 Word Cloud*

1. Apr 21, 23 Text mining, Ch 14-15, Lab 22,23, HW 6.

*Video: Ch 15 Sentiment Analysis*

1. Apr 28, 30 Interactive/Web Ch 19, Lab 24,25 Q6
2. May 5, 7 Shiney Ch 20
3. May 12 Final Exam 8:00-10:00 am CPS 105

**Schedule Footnotes:**

* This schedule is a guide to the coverage of topics. The instructor reserves the right to alter the presentation schedule as necessary to benefit the class.
* No prior experience in programming or statistics is needed.
* Number of labs, homeworks and quizzes may be adjusted as necessary
* Chapter readings, videos, handouts, and lectures are SEPARATE sources for information.
* Videos are available on the book’s website; videos are normally 15 minutes or less and are to be viewed outside of class time.

## Learning Objectives

At the end of the course, students are expected to understand:

* Communication of results to decision-makers
* Essential concepts and characteristics of data
* Scripting/code development for data management using R and R-Studio
* Principles and practices in data screening, cleaning, linking, and visualizations

At the end of the course, students are expected to be able to:

* Cite examples of data analysis successes and failures for different fields.
* Identify a problem and the data needed for addressing the problem
* Perform basic computational scripting using R and other optional tools
* Transform data through processing, linking, aggregation, summarization, visualization, and searching
* Interpret basic statistical measures
* Organize and manage data at various stages of a project lifecycle
* Determine appropriate techniques for analyzing data

**Distribution of Points**

Midterm :20%

Final :20%

Homeworks :20%

Lab Exercises :20%

Quizzes :15%

Attendance/Participation :05% (Includes Lecture and Lab attendance)

**POLICIES**

**Academic Standards -** UW-Stevens Point values a safe, honest, respectful, and inviting learning environment. In order to ensure that each student has the opportunity to succeed, we have developed a set of expectations for all students and instructors. This set of expectations is known as the Community Rights and Responsibilities document, and it is intended to help establish a positive living and learning environment at UWSP. Click here for more information: <http://www.uwsp.edu/dos/Pages/AcademicMisconduct.aspx> Academic integrity is central to the mission of higher education in general and UWSP in particular. Academic dishonesty (cheating, plagiarism, etc.) is taken very seriously. Don’t do it! The minimum penalty for a violation of academic integrity is a failure (zero) for the assignment. For more information, see the “Student Academic Standards and Disciplinary Procedures” section of the Community Rights and Responsibilities document, UWSP Chapter 14. This can be accessed at: [http://www.uwsp.edu/dos/Documents/CommunityRights.pdf - page=11](http://www.uwsp.edu/dos/Documents/CommunityRights.pdf%20-%20page=11)

**ADA Statement** - The Americans with Disabilities Act (ADA) is a federal law requiring educational institutions to provide reasonable accommodations for students with disabilities. For more information about UWSP’s policies, check here: <http://www.uwsp.edu/disability/Pages/faculty/lawAndPolicy.aspx>. If you have a disability and require classroom and/or exam accommodations, please register with the Disability and Assistive Technology Center at the beginning of the course and then contact me. I am happy to help in any way that I can. For more information, please visit the Disability and Assistive Technology Center, located on the 6th floor of the Learning Resource Center (the Library). You can also find more information here: <http://www.uwsp.edu/disability/Pages/default.aspx>.

**Attendance/Participation Policy -** Note: lab assignments indicate attendance. I rarely lecture “STRAIGHT FROM” the book. You are also required to make a 15 minute appointment with me early in the semester so I can get to know you and your expectations for the course.

**Average Time Investment/Workload Policy Statement**

DAC 101 meets twice a week; each meeting is 75 minutes or 2.5 hours per week or 37.5 hours per semester. Additionally, you should expect to spend at least another 5 hours per week, on average, on outside class work including videos and chapter reading assignments.

**Audio/Visual Recording Policy -** Electronic recording of lectures (taping) is prohibited unless receiving prior written approval from the instructor. Approval will be granted only for self-study purposes. You are permitted to take pictures of whiteboards, blackboards or screens of lecture materials, if need be.

**Classroom conduct** – Please mute cell phones and any audible device during classes. Please do not hold private conversations or text while I am lecturing as it is a distraction. Texting or other disruptive activities may lead to points deducted on attendance/participation.

**Canvas –**

* Recorded grades as well as lecture materials (syllabus, PowerPoint class outlines, etc.) will be available on our course 101 Canvas course site.
* It is your responsibility to check that your grades are posted correctly on Canvas. Questions about any posted grade must be raised within TWO weeks of posting. Beyond this time frame, all grade postings are considered correct and final. The Canvas site is NOT available after the final exam.
* USE the OneDrive to save your files – if need be.

**Canvas** –

* **Announcements** on Canvas are the main communication tool (not email!)

**Drop Policy -** In accordance with the rules stated by the College of Professional Studies. I will **NOT** personally drop a student - you are responsible for filling out all the forms.

**Exam Policy -** Except for documented emergencies, no late or makeup in-class exercises, exams and quizzes will be given.

**Grade Policy -** The following scale can always be used to estimate your grade

Percentage breakdown for semester grades (weighted point totals)

A = 93-100% B- = 80-82.99% D+ = 67-69.99%

A- = 90-92.99% C+ = 77-79.99% D = 63-66.99%

B+ = 87-89.99% C = 73-76.99% D- = 60-62.99%

B = 83-86.99% C- = 70-72.99% F = < 60%

\*Instructor reserves the right to implement a curve which is beneficial to the students.

**Homework/Labs Policy** - For lab assignments, you should turn in a single Word document which lists your code, console output and all graphs. If you created or modified a file for use with your program/script, those files need to be turned in as well. If the program is an interactive one, where the user is prompted for input, one or more screen shots (or relevant copy/pastes) of the program’s output are needed to demonstrate the program works correctly. The homework policy is similar to the lab policy. There may be occasions where I ask for more items. If there is any doubt about whether an item should be turned in, err on the side of uploading it to Canvas. I will no longer allow missed items to be turned in later. Finally, when submitting multiple files, do not use zip formats as those files must be downloaded and can not be previewed in Canvas. Failure to abide by these requirements can result in a significant loss of points.

**Homework Policy** – Homework assignments are listed on the syllabus as HW; they are usually due two weeks after the assigned date and are posted on Canvas. Electronic copies of the completed homeworks must be uploaded on the Canvas Assignments by the required date/time - this will be strictly enforced. Only the latest copy of the homework is graded – in case you uploaded multiple copies. Note: You have OneDrive account to save files to if you need to access them later; alternatively, you can save materials on a flash drive but remember to take it with you. Please note that the book’s web site lists labs and homeworks but ALWAYS use what is posted on Canvas. Late assignments are not accepted unless you have some sort of documented emergency; the professor determines if the situation constitutes an emergency.

**Labs** – usually have in-class exercises. Lab time may be redirected as lecture time at the discretion of the instructor.

**Lecture Notes** – electronic version of the notes is available for many topics, however, I strongly encourage you to take good notes as that has been shown to reinforce memory recall.

**Announcements** – Always check the news item on Canvas to find the latest announcements concerning the class.

**PC** – this class will make extensive use of Windows PCs in the lab. If you are not familiar with a desktop computer, you may need to practice a bit with File Explorer and other common utility programs. Note: there are Mac versions of R as well.

**Programming** – technically this is NOT a programming course but you will learn SOME aspects of programming. You will mostly be writing scripts which are just a collection of related program statements. Other courses such as DAC 111, DAC 205 and CNMT 110 are more programming oriented. Keep in mind that data analytics is the intersection of three skill sets: analysis through math/statistics, programming and subject matter expertise which for us is business administration (it could be something else if you are majoring in another field.)

**Software** – Lab Virtual Desktop or install RStudio on your PC, we will use Excel/Access for some assignments. I can help you with the R install on your laptop. 8 gig of RAM should be sufficient for this course but 16 gig is becoming more of the norm and a requirement of DAC 205. By default, R requires all data to be loaded in primary storage.

**Plagiarism Policy -** All assignments and tests should represent YOUR work otherwise you will not receive any credit for that portion of your grade. Disciplinary actions will be pursued for serious offenses.

**Quiz Policy** – quizzes are meant to test your understanding about topics that were currently presented. Quizzes will be take-home but you are NOT allowed to collaborate with others. You may use other resources such as google. For open-ended questions, be careful not to plagiarize. For Quiz questions, make sure to place your answer directly below the question text. Quality of answers matters and is positively correlated with the length of the answer.

**University Emergency Preparedness** –

In the event of a medical emergency call 9-1-1 or use Red Emergency Phones. Offer assistance if trained and willing to do so. Guide emergency responders to victims.

In the event of a tornado warning, proceed to the lowest level interior room without window exposure. See [www.uwsp.edu/rmgt/Pages/em/procedures/other/floor-plans.aspx](http://www.uwsp.edu/rmgt/Pages/em/procedures/other/floor-plans.aspx) for floor plans showing severe weather shelters on campus. Avoid widespan structures (gyms, pools or large classrooms.)

In the event of a fire alarm, evacuate the building in a calm manner. Stay 200 yards away from the building. Notify instructor or emergency command personnel of any missing individuals.

Active Shooter – RUN/ESCAPE, hide, fight. If trapped hide, lock doors, turn off lights, spread out and remain quiet. Call 9-1-1 when it is safe to do so. Follow the instructions of emergency responders.

See UW-Stevens Point Emergency plan at [https://www.uwsp.edu/rmgt](https://www3.uwsp.edu/rmgt)