University of Wisconsin - Stevens Point

College of Letters and Science

Department of Computing and New Media Technologies

Fall 2019 - Version 1.15

Course: Principles of Data and Modeling (DAC 205) #82304

Books: The Art of R Programming (ISBN-13: 978-1786466457)

 BIG DATA ANALYTICS WITH R (ISBN-13: 978-1593273842)

 R for Data Science (Import, Tidy, Transform, Visualize, and Model Data)

 Authors: Grolemund and Wickham

 Available for **free** on http://r4ds.had.co.nz/

 If you'd rather own a paperback copy, it may be ordered (ISBN-13: 978-1491910399)

Class Time: M/W 1:00-2:50 pm (Room SCI B228)

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

Office: CPS 442

Office hours: T: 1-2 pm, W: 11-12 pm & 3-4 pm and by appointment

Contact: kpflugho@uwsp.edu

**Course Description:** Introduction to the concept of data and the multiple perspectives of data. Analyze the data-life cycle from elicitation to disposal, including the steps of collection, processing, storage, use, and archiving. You’ll learn how to get your data into R, get it into the most useful structure, transform it and visualize it.

.

Week # Dates Topic\* Reading/Assignments

1. Sep 4 R & R Studio, Intro ARP1, RDS1, RS1\*,Lab 1
2. Sep 9, 11 Data Categories, Vectors, Matrices ARP2, RDS2, RS2\*,Lab 2, HW1
3. Sep 16, 18 Data Frames, Tibbles, Data Cat. ARP3&5, RDS3, RS3\*, Lab 3, Q1
4. Sep 23, 25 tidyVerse, Functions, VM ARP4, Lab 4, HW2
5. Sep 30, Oct 2 Relational Data, Linux ARP6, RDS4, Lab 5,Q2

 ***OCT 4 GREAT LAKES ANALYTICS CONFERENCE***

1. Oct 7, 9 ggPlot2 - Visualization ARP12, RDS5, Lab 6, HW3
2. Oct 14, 16 Strings & Regular Expressions ARP11, RDS6, Lab 7,Q3
3. Oct 21, 23# Review, #Midterm Exam .
4. Oct 28, 30 MetaData, Missing Values, Wgts ARP7, BDA1, Lab 8
5. Nov 4, 6 Big Data, DPlyr, Sparklyr ARP10, BDA 3, Lab 9, Q4, HW4
6. Nov 11, 13 Parallelism Lab 10, BDA 4 (pp 127-191)
7. Nov 18, 20 Hadoop Lab 11, Q5, HW 5, BDA 7
8. Nov 25, 27 Spark with R Lab 12, BDA 8
9. Dec 2, 4 More Spark Lab 13,Q6, HW6, BDA 9
10. Dec 9, 11 Catchup/Review Lab 14
11. Dec 17 Final Exam 12:30-2:30 pm SCI B228

**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.
* RS Videos are from the RStudio Webinars
* Chapter readings, handouts, and lectures are SEPARATE sources for information.

**Course Outcomes** - Given a successful conclusion of this course, students will be able to:

* Understand fundamental data types and structures
* Be able to process data using appropriate structures
* Understand the importance of metadata and data stewards
* Prepare incomplete and/or big data for modeling
* Know different approaches to handle structured versus unstructured data
* Understand data life cycle issues including data security
* Apply many of the above outcomes via working R programs
* Learn about data and how to input, store, and output data in R
* Know basic Linux commands
* Operate between a host Windows machine and a Virtual Linux machine
* Learn how to process big data via the VM
* Examine Hadoop and related technologies

**Distribution of Points**

Midterm :15%

Final :20%

Labs :20%

Quizzes: :15%

Homeworks :25 %

Attendance :05% (Includes Lecture and Lab attendance)

R: The open source R package provides a complete analytical environment and is justifiably viewed as the de facto language in the academic statistical community. Its popularity in the non-academic, business-orientated world is growing rapidly as well. It is in fact the fastest growing language on the StackOverflow developers

site and is currently ranked in 5th place in IEEE language rankings. We will begin with R programming basics and will assume no prior R knowledge. The use of the R Integrated Development Environment (IDE), RStudio (Desktop Open Source License), will be emphasized throughout the semester.

Videos:

RStudio Webinars Series (https://www.rstudio.com/resources/webinars/) for learning R and RStudio

|  |  |
| --- | --- |
| RS1 |  Programming Part 1 (Writing code in RStudio) |
| RS2 |  Programming Part 2 (Debugging code in RStudio) |
| RS3 |  Importing Data into R |

Readings: The Art of R Programming

|  |  |
| --- | --- |
| ARP 1 | Getting Started |
| ARP 2 | Vectors |
| ARP 3 | Matrices and Arrays - includes some image processing |
| ARP 4 | Lists |
| ARP 5 | Data Frames |
| ARP 6 | Factors and Tables |
| ARP 7 | R Programming Structures |
| ARP 10 | Input/Output |
| ARP 11 | Strings |
| ARP 12 | Graphics |
| ARP 13 | Debugging  |

Readings: R for Data Science (online)

|  |  |
| --- | --- |
| RDS 1 | Welcome: Introduction, Explore: Introduction |
| RDS 2 | Explore: Workflow Basics |
| RDS 3 | Wrangle: Tibbles |
| RDS 4 | Wrangle: Relational Data |
| RDS 5 | Explore: Data Visualization |
| RDS 6 | Wrangle Strings |

Readings: BIG DATA ANALYTICS WITH R

|  |  |
| --- | --- |
| BDA 1 | The Era of Big Data |
| BDA 3 | Unleashing the Power of R from within |
| BDA 4 | Hadoop and MapReduce Framework |
| BDA 7 | Faster than Hadoop - Spark with R |
| BDA 8 | Machine Learning Methods for Big Data in R |
| BDA 9 | The Future of R - Big, Fast and R Data |

 **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%3D11)

**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 Policy -** Attendance will be taken randomly in lecture/lab and will count towards your grade! I rarely lecture “STRAIGHT FROM” the book.

**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 allowed to take pictures of whiteboards, blackboards or screens of my lecture material, if need be.

**Average Time Investment/Workload Policy Statement**

DAC 205 meets twice a week; each meeting is 110 minutes or about 4 hours per week or 64 hours per semester. Additionally, you should expect to spend at least another 8 hours per week, on average, on outside class work including videos and chapter reading assignments.

**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. No FOOD or DRINKS are allowed in the lab.

**Canvas –**

* Recorded grades as well as lecture materials (syllabus, PowerPoint class outlines, etc.) will be available on our course 205 Canvas course site - except material written on whiteboards or typed during class.
* 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.
* **Announcements** on Canvas is the main communication tool (not email!)

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

**Email Policy**

* I try to answer questions in a timely manner but if you haven’t received a response from me by the end of the next business day, please resend the email.
* If your email is only informative in nature, such as you are missing a class, I usually don’t reply to those emails but rather just file them. If your email has a question or issue that needs to be addressed, I will reply to it.
* Please include “DAC 205” as part of your subject line.

**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 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. Format for the electronic file is listed in the lab bullet. Late assignments are not accepted. 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. 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. For lab assignments, you should turn in a Word document which lists your code, and one or more screen shots (or relevant copy/pastes) of the program’s output to demonstrate the program works correctly. If you have extra lab time, you are encouraged to work on your DC assignments. Number of lab assignments may vary from schedule.

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

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

**Software** – Lab Virtual Desktop or install RStudio on your PC, we may use Excel for some assignments. I can help you with the R install on your laptop. Although 8 gig of RAM should be sufficient R and R Studio, this class will take advantage of a virtual machine. The virtual machine may require 16 gig of RAM.

For laptop installs:

Download R at https://cran.r-project.org/

Download RStudio: https://www.rstudio.com/products/rstudio/download/#download

**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 – see Academic Standards.

**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.

**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)