

University of Wisconsin - Stevens Point
 College of Letters and Science
 Department of Computing and New Media Technologies
 Fall 2018 - Version 1.11

Course: Principles of Data (DAC 205) #83311
 Books: The Art of R Programming (ISBN-13: 978-1593273842) – Purchase at bookstore
 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 11:00-12:00 pm (Room CPS 208)
 M 12:00-12:50 pm (Room CPS 105)
 Professor: Dr. Kurt A. Pflughoeft (Floog'heft)
 Office: Science B231
 Office hours: M 2-3 pm, W 12-1 pm, TR 11-12 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.

<u>Week #</u>	<u>Dates</u>	<u>Topic*</u>	<u>Reading/Assignments</u>
1.		R & R Studio	ARP1, RDS1, DC1, RS1*
2.	Sep 10	Data Categories, Vectors, Matrices	ARP2, RDS2, DC2, RS2*
3.	Sep 17	Data Frames, Tibbles, Data Cat.	ARP3&5, RDS3, DC3, RS3*, Lab 1, Q1
4.	Sep 24	tidyVerse, Functions	ARP4, DC4, Lab 2
5.	Oct 1	Relational Data	ARP6, RDS4, DC5, Lab 3, Q2
6.	Oct 8	ggPlot2 - Visualization	ARP12, RDS5, DC6, Lab 4
7.	Oct 15	Strings & Regular Expressions	ARP11, RDS6, DC7, Lab 5, Q3
8.	Oct 22#	Review, #Midterm Exam	.
9.	Oct 29	Data Quality/Life Cycle	ARP7, DC8, Lab 6
10.	Nov 5	MetaData, Missing Values, Wgts	ARP10, Lab 7, Q4
11.	Nov 12	Data Anonymization	DC9, Lab 8.
12.	Nov 19	Text Mining - Bag of Words	DC10, Lab 9, Q5
13.	Nov 26	TM - Sentiment/Fuzzy Matching	DC11, Lab 10
14.	Dec 3	Data Security	Lab 11, Q6
15.	Dec 10	Big Data and R	DC12, Lab 12
16.	Dec 18	Final Exam 10:15-12:15 pm	CPS208

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 optional (RStudio Webinars)
- The second half of class will be a lab
- 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

Distribution of Points

Midterm	:15%
Final	:20%
Labs	:25%
Quizzes:	:10%
DataCamps	: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.

DataCamp/Homeworks: Students will use DataCamp (www.datacamp.com), an online training resource for learning data programming in R. To access the materials on DataCamp, students must join the course group via an email invite. Each student will have a free 6-month subscription when signing into our class group. Find our class group at: <https://www.datacamp.com/enterprise/dac205> DataCamp will be used for instructional videos as well as integrated homework assignments. I will transfer your DC points to D2L after the due date. Please note that if hints are accepted within Data Camp that will reduce the total amount of points one can earn. Note: To hear the audio for videos in the computer labs requires ear buds.

Data Camp

DC1	Introduction to R: Intro to Basics, Vectors
DC2	Introduction to R: Matrices, Factors
DC3	Introduction to R: Frames and Lists
DC4	Intermediate R: Functions
DC5	Introduction to TidyVerse: Data Wrangling, Data Visualization
DC6	Introduction to TidyVerse: Groups & Summarize, Types of Visualizations
DC7	String Manipulation in R with stringr: String Basics, Intro to Stringr
DC8	String Manipulation in R with stringr: Pattern matching with regular expressions, More advanced matching and manipulation
DC9	Data Privacy and Anonymization in R: Intro to Data Privacy
DC10	Text Mining: Bag of Words: Jumping into..., Word Clouds and ...
DC11	Text Mining: Bag of Words: Adding to your TM Skills, Battle of the tech giants for talent
DC12	Scalable Data Processing in R: Working with increasingly large data sets, Processing and Analyzing Data with bigmemory

The professor may alter the above Data Camp chapters to benefit the class.

Optional 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 – OPTIONAL

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

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

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 once a week; each meeting is 110 minutes or about 2 hours per week or 32 hours per semester. Additionally, you should expect to spend at least another 4 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.

D2L –

- Recorded grades as well as lecture materials (syllabus, PowerPoint class outlines, etc.) will be available on our course 205 D2L course site.
- It is your responsibility to check that your grades are posted correctly on D2L. 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 D2L site is not available after the final exam.
- It's best to download attachments rather than view directly in the browser.
- USE the OneDrive to temporarily save your files – if need be.
- **News feed** on D2L is the main communication tool (not email!) - consider the text msg option
- The UW system is in the process of moving to a different LMS called Canvas

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.

Extra Credit (Max Total 6%): You will have the opportunity to complete extra credit assignments. Extra credit assignments are optional. Your extra credit opportunities are to complete a R-related DataCamp which has not been assigned as part of this course. Each DataCamp course is worth up to 2 standardized percentage points; you can do a max of three additional courses for extra credit.

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.

Labs –are scheduled from 12:00 – 12:50 p.m. on Mondays and 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.

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 news item on D2L to find the latest announcements 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. 8 gig of RAM should be sufficient for an academic environment but data analysts in industry often require more as all data must be loaded in primary storage for R.

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 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>