Data II Data Wrangling
TERM
FACULTY
University of Kansas

Contents

1 Preliminary Details

1.1 Course Information

Course Number: DS3XX
Course Name: Data Wrangling
Location: TBD
Time: TBD
Start/End Date: TBD
Final Exam Time: TBD

1.2 Professor Information

Professor: TBD
Office: TBD
Office Hours: TBD
Email: someone@ku.edu

2 Course Description, Objectives and Textbook

2.1 Course Objectives

Course Description:
Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to derive knowledge and insights from data. This course teaches students the critical concepts of inference and computing, working with real behavioral, economic, geographic, physical, social, and text data. Students obtain basic statistics training from a computational perspective using simulation to answer questions, explore problems, and delve into social issues surrounding data analysis such as privacy and design.

Course Objectives/Student Learning Outcomes:
Upon successful completion of this course, students will be able to:

- Developed an appreciation for the value of data science in psychology
- Gained proficiency and confidence in foundational data science skills
- Applied these skills to extract knowledge and insights from real-world data
- Laid the groundwork for creating a data science portfolio, blog, or website
- Been exposed to advanced data science topics that piqued their interest
- Pushed through enough uncertainty to feel empowered to continue learning
- Begun accepting that silly mistakes are inevitable and need to be planned for
- Automated a tedious task that no human will ever have to do manually again

- Benefited from the generosity of others and paid it forward to someone else
- Felt like a total fool, an absolute wizard, and just about everything in between
- Found their inner child, pursued some whimsies, and had fun along the way

Prerequisite: Data II

2.2 Textbook

Most of the material that students will be required to learn will come from the course lectures, activities, and assignments. However, students' learning will also be supplemented by assigned readings from the following textbook. It is available in physical/printed form and you are welcome to purchase a copy if you prefer printed books. However, a web-based version is also available for free and it is recommended you try this version as it is more frequently updated, easier to navigate (without pagination), and allows you to easily follow along by copying and pasting its examples.

- R for Data Science by Hadley Wickham & Garrett Grolemund

The following texts are where I would recommend you turn after completing this course (or during it, if you are feeling ambitious) to dive into more advanced topics about visualization and programming, respectively.

- ggplot2: Elegent graphics for data analysis (3rd Edition) by Hadley Wickham
- Advanced R (2nd Edition) by Hadley Wickham

2.3 Technology

We will use computational and statistical programming during this class. The softare package used for labs and assignments is R/RStudio.

2.4 Grading

2.5 Grading Scheme

There are 6 components of the class (described below). Each component is worth 100 points. Your final point total is a weighted average of each of these components. It is out of 100 points and is rounded to the nearest integer. The final point total is set according to the following formula:

$$\begin{aligned} \text{Final Grade} &= .05 \times \text{Participation} \\ &+ .20 \times \text{Homework} \\ &+ .15 \times \text{Lab Assignment} \\ &+ .15 \times \text{Low Exam} \\ &+ .20 \times \text{High Exam} \\ &+ .25 \times \text{Final Exam} \end{aligned}$$

Your final grade is assigned according to the following break down:

Lower Real Limit	92.5	89.5	86.5	82.5	79.5	76.5	72.5	69.5	66.5	62.5	59.5	0
Grade	4.0	3.7	3.3	3.0	2.7	2.3	2.0	1.7	1.3	1.0	0.7	0

2.6 Class Participation

Everybody will be expected to come to class having read the relevant readings and being prepared to participate in class. Reading assignments are listed in the course schedule below. Complete each reading assignment before the date for which it is listed. Participation points must be earned by attending class meetings. We will calculate the percentage of class meetings that you attend. Course participation grade is assigned according to the following rule

Lower Real Limits	79.5	74.5	69.5	64.5	59.5	64.5	0
Points	100	90	80	70	60	50	0

So a student who attends 80% of the meetings will receive 100 points. A student who attends 79% of the meetings will receive 90 points.

2.7 Homework

Weekly homework assignments are a required part of the course. Each student must submit each homework independently, but you are allowed to discuss problems with other students and course staff.

The homework will use computational notebooks and the python programming language. You do not need prior experience with these tools to be successful. These tools will be introduced in class.

You can turn in your homework assignments 5 days late with a 20% deduction. After 5 days, we will not accept late submissions because we will release answer keys by then. There will be approximately ten homework assignments. The lowest scored homework will be dropped.

2.8 Lab Assignments

Weekly labs are a required part of the course. After the first week, labs will be released on Sunday night. You can get credit for each lab in one of two ways described below:

- Attend class, make progress substantial enough for your work to be checked off by course staff, and submit your lab (even if it is incomplete) by the end of the class period on the day the lab is due. You need to attend both class meetings of the designated week to get full credit this way.
- Complete the lab independently and submit the completed lab by end of class the day it is due (often Wednesday). "Completing a lab" means passing all tests included in the lab.

There will be approximately 10 labs. The lowest scored lab will be dropped.

2.9 Exams

There are two exams in the class during the semester. These exams contain an in-class portion and a take-home portion. There is also a final exam. This exam has only an in-class portion.

Unless you have accommodations as determined by the university and approved by the instructor, you must take the exams at the dates and times provided here.

2.10 Grade Appeal

Grades on assignments are final. If you feel there is an error in your grade, you must submit a request to me in writing to re-evaluate your assignment. If the error is arithmetic, your grade will be updated without further action. If the error is of another kind, the professor reserves the right to re-grade the entire assignment.

3 Course Policies

3.1 Course Structure

The class will be taught in-person as a flipped course. Students are expected to watch the lectures and read the material before coming to class. Lab work will be done in class, and there should be time to work on some of the homework.
It is important that you spend the necessary time working in this course to achieve the expected outcomes by the end of the term. If you face challenges to fully participating at any time during the semester, please let me know, and please contact me if you expect to miss class. I am available and ready to support your success. Additionally, if you need to report an extended illness or serious accident, please contact Student Support and Case Management at course-adapt@ku.edu or 785-864-4060. A case manager will send email notifications to your instructor(s) on your behalf. When you are able, you will need to follow up with your professors to coordinate a temporary arrangement regarding missed instruction and coursework.

3.2 Policy for Late Assignments

You can turn in your assignment five days late with a 20% deduction. After 5 days, we will not accept late submissions because we will release answer keys in a week.

4 Academic Misconduct

The University community traditionally has been a place where all members are free to express and exchange ideas. Such fundamental goals of the University as intellectual growth and development are predicated on honest investigation, straightforward expression of views and opinions, and genuine dialogue. The attainment of these goals requires that all who participate in the exchange of ideas maintain intellectual integrity. The University of Kansas seeks to ensure that both the professor and the student are protected from unfair actions or accusations in cases of cheating and plagiarism. The University encourages faculty and students to adopt a responsible attitude toward one another.

Academic misconduct includes but is not confined to plagiarizing; cheating on assignment and assessments; turning in counterfeit reports, tests, and papers; stealing tests and other academic material; knowingly falsifying academic records or documents; and turning in the same work to more than one class. Students and faculty alike must recognize that none of the procedures set forth in this document operate to the exclusion of civil or criminal litigation. Likewise, no definitions in this document supersede any parties concerned to resolve the contested issues without the necessity for recourse to the law in a manner that protects the rights of the individuals involved.

Consequences of academic misconduct may include, but are not limited to, a failing grade for an assignment, a failing grade for a course, or expulsion from the University. Any form of academic misconduct which results in administrative or academic withdrawal is noted on the student's transcript.

Collaboration Policy: Discussion and the exchange of ideas are essential to academic work. You are encouraged to consult with your classmates on the homework assignments and to share sources. However, you should ensure that any work you submit for evaluation is the result of your own research and that it reflects your own approach to the topic. No collaboration of any kind is allowed during the midterms or final exam.

5 Credit Hour Definition

Consistent with best practices in higher education, the University of Kansas subscribes to the federal definition of the "credit hour" endorsed by the Higher Learning Commission. Driven by intended learning outcomes and verified by evidence of student achievement, the "credit hour" is an institutionally-established equivalency that reasonably approximates not less than one hour of classroom (or direct faculty) instruction and a minimum of two hours of out-of-class student work per week for the duration of the course enrollment period. A 3-credit-hour course, for example, requires approximately 40-45 instruction (or instructional equivalency) hours, roughly 80-90 out-of-class work (self-directed) hours and approximately 120-135 total instructional hours over the course of the semester.

6 ADA Statement

TheAcademic Achievement and Access Center (AAAC) coordinates academic accommodations and services for all eligible KU students with disabilities. If you have a disability for which you wish to request accommodations and have not contacted the AAAC, please do so as soon as possible. They are located in 22 Strong Hall and can be reached at 785-864-4064 (V/TTY). Information about their services can be found at http://www.disability.ku.edu. Please contact me privately concerning your needs in this course.

7 Course Schedule

NOTE: the course schedule is subject to change at professor's discretion.

Day	Concept/Activity	Reading	HW Due
Week 1			
8/22 Tuesday	Introduction and Syllabus	Chapter 1	
8/24 Thursday	Meeting R		
(Due Tuesday)	**Lab 01**		
Week 2			
8/29 Tuesday	Basic Programming		
8/31 Thursday	More Programming		
(Due Tuesday)	**Lab 02**		
Week 3			
9/5 Tuesday	Tidy Data		
9/7 Thursday	dplyr Basics		
(Due Thursday)	**Lab 03**		
Week 4			
9/12 Tuesday	ggplot basics		
9/14 Thursday	More ggplot		
Week 5			
9/19 Tuesday	Program to Communicate		
9/21 Thursday	Visualize to Communicate		
(Due Tuesday)	**Lab 04**		
Week 6			
9/26 Tuesday	Unit I Project		
9/28 Thursday	Unit I Project		
(Due Tuesday)	**Lab 05**		
Week 7			
10/3 Tuesday	Unit I Project		
10/5 Thursday	Unit I Project		
(Due Tuesday)	**Lab 06**		
Week 8			
10/10 Tuesday	In Class Exam I		
10/12 Thursday	FALL BREAK		

Day	Concept/Activity	Reading	HW Due
Week 9			
10/17 Tuesday			
10/19 Thursday			
(Due Tuesday)	**Lab 07**		
Week 10			
10/24 Tuesday	Estimation and confidence	Chapter 13	
10/26 Thursday	Center, Spread, and Normal Distribution	Sections 14.1-14.3	HW07: Testing Hypotheses (11/3)
(Due Tuesday)	**Lab 08**		
Week 11			
10/31 Tuesday	Normal Distribution and Variance of Sample Means	Sections 14.4-14.6	
11/2 Thursday	Sampling Variability		HW08: Confidence Intervals (11/10)
Week 12			
11/7 Tuesday	Correlation	Sections 15.1-15.2	
11/9 Thursday	Regression	Sections 15.3-15.6	HW09: Bootstrap, Resampling, CLT (11/17)
(Due Tuesday)	**Lab 09**		
Week 13			
11/14 Tuesday	Regression Inferences	Chapter 16	
11/16 Thursday	Classification	Chapter 17	HW10: Regression (11/28)
(Due Tuesday)	**Lab 10**		
Week 14			
11/21 Tuesday	NO CLASS		
11/23 Thursday	THANKSGIVING BREAK		
Week 15			
11/28 Tuesday	Review		
11/30 Thursday	In Class Exam II		
Week 16			
12/5 Tuesday	Take Home Exam II Due		
12/7 Thursday	Review		
Finals Week			