DA 101: Introduction to Data Analytics

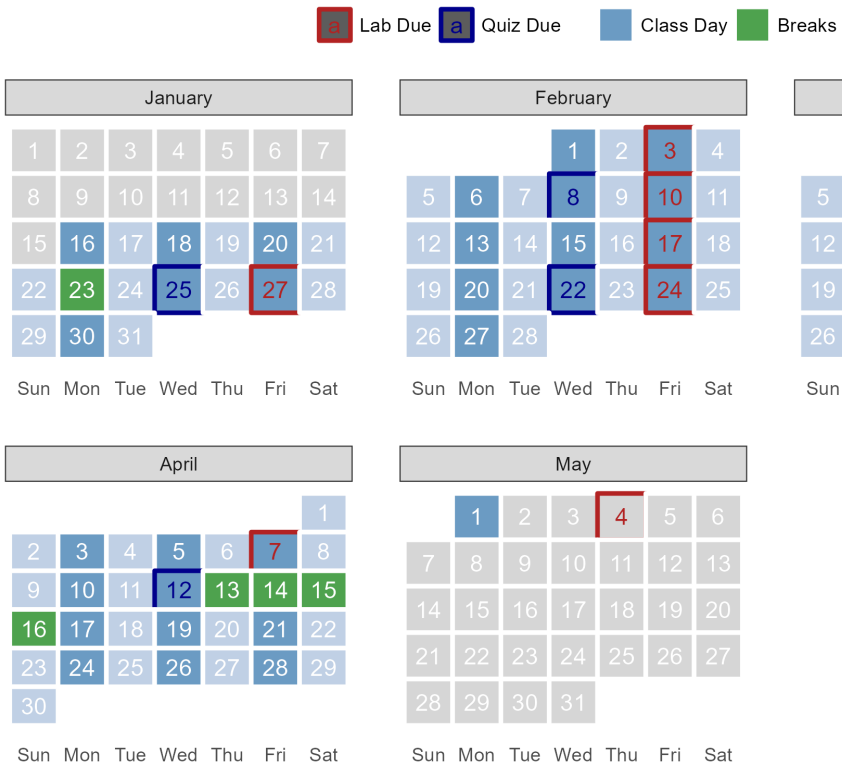
Dr. Miles D. Williams

Spring 2023

E-mail: williamsmd@dension.edu
Office Hours: MW 3:00-4:30 pm
Office: Higley 401C

Web: milesdwilliams15.github.io
Class Hours: MWF 11:30 am-12:20 pm | Lab: F 1:30-4:20 pm
Class Room: Burton D. Morgan Center 219

Course Outline



Notable dates: MLK Day Observed (23 Jan.), Spring Break (11-20 Mar.), MPSA (13-16 April), Final Day of Instruction (01 May)

Basic set-up for week

- Day 1: Topics
- Day 2: Topics + quiz (every other week)
- Day 3: Discussion + Lab Day

Note: All readings for discussion and other course materials will be available on Canvas. In-class exercises will involve real-time feedback using this form.

Part I: Getting Started with Data and Coding

Week 01, 01/16 - 01/20

- Day 1: Course Intro
- Day 2: Data types, entry, tidying, and quality control and assurance
- Day 3: Data in your world
- Lab: Invasive Species I

Week 02, 01/23 - 01/27

- Day 1: MLK Observed [NO CLASS]
- Day 2: Metadata and data analysis cycle — *Quiz due Wed. by midnight*
- Day 3: Open Analytics, Equity, Diversity, and Inclusion
- Lab: Data Collection — *Invasive Species I due Fri. by 1:30 pm*

Week 03, 01/30 - 02/03

- Day 1: Data and Databases
- Day 2: Wrangling and dplyr
- Day 3: 1-on-1s with me
- Lab: Invasive Species II — *Data Collection due Fri. by 1:30 pm*

Week 04, 02/06 - 02/10

- Day 1: Data viz
- Day 2: Data viz — *Quiz due Wed. by midnight*
- Day 3: Challenges with missing data
- Lab: Code Glossary I — *Invasive Species II due Fri. by 1:30 pm*

Part II: Answering Questions with Statistical Analysis

Week 05, 02/13 - 02/17

- Day 1: Description
- Day 2: Hypothesis tests
- Day 3: Data Visuals and Performing Arts
- Lab: Substance Use — *Code Glossary I due Fri. by 1:30 pm*

Week 06, 02/20 - 02/24

- Day 1: Prediction
- Day 2: Linear regression — *Quiz due Wed. by midnight*
- Day 3: DA in social and political science
- Lab: Political Polarization I — *Substance Use due Fri. by 1:30 pm*

Week 07, 02/27 - 03/03

- Day 1: Correlation
- Day 2: Causation
- Day 3: Statistical ethics
- Lab: Political polarization II — *Political Polarization I due Fr. by 1:30 pm*

Part III: It's Complicated

Week 08, 03/06 - 03/10

- Day 1: Ethics & Bias
- Day 2: Ethics & Bias — *Quiz due Wed. by midnight*
- Day 3: Analysis case study - quantitative research
- Lab: Code Glossary II + Peer grading activity — *Political Polarization II due by 1:30 pm*

Week 09, 03/13 - 03/17

SPRING BREAK!

Week 10, 03/20 - 03/24

- Day 1: Multivariate regression
- Day 2: Model selection
- Day 3: Data analysis in a global pandemic
- Lab: Airbnb in Columbus — *Code Glossary II due Fri. by 1:30 pm*

Week 11, 03/27 - 03/31

- Day 1: Maps + spacial data
- Day 2: Maps + spacial data — *Quiz due Wed. by midnight*
- Day 3: Human data ethics workshop
- Lab: IRB cert + Ethical Frameworks — *Arbnb in Columbus due Fri. by 1:30 pm*

Week 12, 04/03 - 04/07

- Day 1: Write your own functions + style guides
- Day 2: Write your own functions + style guides
- Day 3: Debuging + code reading activity
- Lab: Code glossary III + introduce final project — *Ethical Frameworks due Fri. by 1:30 pm*

Part IV: Choosing your own path and gaining independence

Week 13, 04/10 - 04/14

- Day 1: Regular expressions and community-building
- Day 2: Share initial questions, codebook — *Quiz due Wed. by midnight*
- Day 3: I'm at MPSA so NO CLASS
- Lab: I'm at MPSA so NO LAB

Week 14, 04/17 - 04/21

- Day 1: Open analytics + debuging
- Day 2: Shiny
- Day 3: Progress report
- Lab: Project sprints

Week 15, 04/24 - 04/28

- Day 1: Presenting data analysis
- Day 2: Student presentations and peer review
- Day 3: Next gen data scientists
- Lab: Presentations + report

Week 16, 05/01 - 05/05

- Day 1: Student presentations + peer review *Final Project due Thursday, 05.04 by 11:00 am*

Course Description

Data surround us everywhere we look. But where does it come from? What can we learn from it? A wide range of processes—politics, business, sports, health, lab and field experiments, physical phenomena—generate data. In this class you will be introduced to various kinds of data and the tools used to wrangle and analyze it to answer questions and gain insights. *No previous experience with data analytics or statistical programming is required*.

Class on Mondays and Wednesdays will consist of short introductions to new concepts followed by hands-on computing exercises. On Fridays, we'll focus on more abstract and conceptual issues related to data-driven puzzles, ethical conundrums, and creative approaches to data analysis. Friday afternoon labs will be dedicated to in-depth practice of the skills you will learn via hands-on exploration of data from a variety of fields. Some labs will consist of individual projects, and others will be done in teams.

Course Objectives

In this course, you'll develop a number of skills.

Quantitative You will develop your quantitative skills in the treatment of data. You'll learn how data are gathered, assembled into datasets, and most effectively analyzed and visualized to describe and draw inferences.

Writing & Communication Even the best analysis requires communication of findings to an audience. You'll get experience presenting your findings both in writing and oral presentations.

Critical Thinking Critical thinking involves "confronting multiple, competing perspectives and adjudicating between them." This is the essence of what is required to organize data and draw conclusions.

Statistical Programming You will use R and RStudio to wrangle and analyze data. By the time you have finished this course you may not be an expert programmer, but you will be well versed in the basics of R and how a variety of R packages to assemble, organize, analyze, and visualize data.

Teamwork Some projects work will be done in teams. In data analytics, teamwork really does make the dream work. Aside from that, it is unavoidable. Most data analysis in the real world is a collaborative exercise, so learning how to work well with others is essential.

Analysis Software & Technology

Students will use R and RStudio, which are free and open source! Follow this guide to install each locally to your laptop. Support will be provided by the instructor in the installation of required software. If at any time you don't have access to a laptop please contact the instructor and the Data Analytics Program can provide you with a loan from the laptop cart.

As we work through learning new code and concepts, it helps me to know what you understand and what you need to spend more time on. To make that happen, in each class you will provide real-time feedback using this form. This way I know when I can introduce something new and when I need to provide more help and examples.

Required Readings

In addition to reading some short pieces written by journalists, academics, and others available online, we will primarily draw on *R for Data Science*, which is both **free** and available **online**.

> Hadley Wickham and Garrett Grolemund. 2017. *R for Data Science: Visualize, Model, Transform, Tidy, and Import Data*. https://r4ds.had.co.nz/

If you wish, you can also order a physical copy from Amazon.

Resources

We Have a TA!

If you need additional help outside of class, we will have a DA 101 TA available to help. The TA will also be present during Friday labs. More details to come by the instructor.

Accessibility

Students with a documented disability should complete a Semester Request for Accommodations through the MyAccommodations app on MyDenison. It is the student's responsibility to contact me privately as soon as possible to discuss specific needs and make arrangements well in advance of an evaluation. I rely on the Academic Resource Center (ARC) located in 020 Higley Hall, to verify the need for reasonable accommodations based on the documentation on file in that office. Reasonable accommodations cannot be applied retroactively and therefore ideally should be enacted early in the semester as they are not automatically carried forward from a previous term and must be requested every semester.

Writing Center

Staffed by student Writing Consultants, the Writing Center is a free resource available to all Denison students. Writing Consultants from a range of majors work with writers one-on-one in all phases of the writing process, including (but not limited to): deciphering assignments, discussing ideas, developing an argument, integrating research and sources, working with faculty feedback, and/or polishing a draft. In addition, Consultants are happy to help with all types of writing, from lab reports, research papers, and informal writing assignments to cover letters, personal statements, and other application materials. The Center welcomes writers from all backgrounds and levels of college preparation, including multilingual writers. Should a multilingual writer need writing assistance that exceeds the abilities of consultants, the writer can be referred to the Coordinator for Multilingual Learning, Kaly Thayer (thayerk@denison.edu). Writing Center consultations will take place in person in in the Atrium level of the Library; please visit the Writing Center's page (https://my.denison.edu/campus-resources/writing-center) on MyDenison for specific information regarding hours of availability and how to schedule an appointment. The Writing Center strongly recommends signing up for appointments in advance.

Multilingual Support (L2)

Students who use English in addition to other languages are welcome to use the resources available at the Multilingual Learning Office. Kaly Thayer, the Assistant Director of Multilingual Learning, and Anna Adams, the English Language Support Specialist, as well the student consultants who work with them, are trained and experienced in helping students address the different issues that arise when working in more than one language. If English is not your first or only language, please consider utilizing this resource, which is available to ALL Denison students. Ms. Thayer, Ms. Adams, and the student consultants offer a variety of support for L2 students, including consulting with you about your written language (grammar, syntax, word-choices), strategies to manage your reading assignments, assistance with class conversation and presentations, and help devising ways to develop and effectively use all your skills in English. You can set up an appointment via MyDenison - Campus Resources - Multilingual Learning, or by emailing the Multilingual Learning Office directly at englishhelp@denison.edu.

Reporting Sexual Assault

Essays, journals, and other coursework submitted for this class are generally considered confidential pursuant to the University's student record policies. However, students should be aware that University employees are required by University policy to report allegations of discrimination based on sex, gender, gender identity, gender expression, sexual orientation or pregnancy to the Title IX Coordinator or a Deputy Title IX Coordinator. This includes reporting all incidents of sexual misconduct, sexual assault and suspected abuse/neglect of a minor. Further, employees are to report these incidents that occur on campus and/or that involve students at Denison University whenever the employee becomes aware of a possible incident in the course of their employment, including via coursework or advising conversations. There are others on campus to whom you may speak in confidence, including clergy and medical staff and counselors at the Wellness Center. More information on Title IX and the University's Policy prohibiting sex discrimination, including sexual harassment, sexual misconduct, stalking and retaliation, including support resources, how to report, and prevention and education efforts, can be found at: https://denison.edu/campus/title-ix .

R Resources & Cheat Sheets

How to Google R Stuff ◊ `{tidyr}` ◊ `{dplyr}` ◊ `{ggplot2}` ◊ `{rmarkdown}` ◊ `{seerrr}` ◊ `{coolorrr}`

Course Policy

The course policy and requirements are detailed below. It all basically boils down to: (1) show up to class, (2) learn some stuff, and (3) don't cheat or trick me into believing you've accomplished 2.

Grading Policy

Grades at Denison are based on a standard 4.0 scale. You can read more about Denison's grading system here. Generally, a 90 corresponds to an A−, an 80 to a B−, etc.

Grading Scale		
A+: 98%+	A: 92%	A-: 90%
B+: 88%	B: 82%	B-: 80%
C+: 78%	C: 72%	C-: 70%
D+: 68%	D: 62%	D-: 60%
F: below 60		

Code Glossary **10pts**

> We're going to be using statistical programming software in this class, which means you're going to need to learn how to code. There's no expectation that you have any coding experience prior to taking this course, but whether you have or haven't coded before, taking notes and building a repository of code that you can draw on in future work is essential. That's why I want you not only to take notes in class as we write code but also to build a code repository that includes examples of different kinds of analysis and how to process the data used to perform them. The good news is, you won't have to come up with these examples on your own. If you code along with me in class, you'll have a complete code glossary for each of the parts of the course. Different code glossaries will be due each time we start a new unit of the course.

Fortnightly Quizzes **5pts Total**[1]

> These will be short 5-10 minuted (timed) quizzes that you will take on Canvas to test your knowledge and retention of skills learned in class and in Lab. You will take these on your own time and will be due every other Wednesday (that is, fortnightly!) by midnight. The lowest grade will be dropped.

Friday Reading Responses **5pts Total**

> Every Friday our normal class time will be dedicated to in-class discussion of some reading. I expect everyone to come to class prepared to participate. That means having notes, coming to class with questions and comments about what we read, and engaging in discussion. As a commitment device for you (and to provide me with evidence that you've done the reading), each Friday before class you will submit a short

[1] I always wanted an excuse to use "fortnightly."

write-up in response to a prompt that I will post to Canvas related to the reading. Tese will be graded on a 0-3 scale. The lowest grade will be dropped.

Friday Lab Projects **60pts Total**

Weekly labs will challenge you to practice, apply, and extend the concepts and skills learned in the course. There will be five topics from different domain areas where data analytics can provide insight into interesting questions, and hopefully, some answers. Each topic will increase in complexity, skills, and expectations. Later labs will be worth more points than earlier labs.

Some labs will be short individual assignments to practice skills introduced that week. The last lab within each topic will be a group project that will require you to collaborate to apply skills from previous weeks to investigate a set of questions. These will conclude with a group report and/or presentation of the results.

Due on Canvas by 1:30 pm on Friday, before the beginning of lab.

Final Project **20pts Total**

The final project will be a culmination of the skills you have learned in the weeks prior, and will offer more flexibility in your design, data choices, and approach, while demonstrating mastery of the breadth of skills discussed in the course. Potential topics will be discussed and approved with the instructor. Final reports will discuss the data, analytical approach, and a visual and oral presentation of the results.

Due on Canvas May 4 by 11:00 am.

$$\begin{aligned} 10pts &\text{ (Code Glossaries)} \\ 5pts &\text{ (Quizzes)} \\ 5pts &\text{ (Reading Responses)} \\ 60pts &\text{ (Lab Projects)} \\ +20pts &\text{ (Final Project)} \\ \hline 100pts &\text{ (Total Assignments)} \end{aligned}$$

Wait! What about attendance? You're all young adults. I expect you to come to class and participate. But I won't be taking attendance (at least not formally). If you can't make it to class for some reason, then you can't make it to class. But if you start missing class regularly, your grade will start to suffer. It won't suffer because you're losing "participation" points. It'll suffer because you're missing out on information provided in class that's necessary for completing Lab work, doing quizzes, and building your code glossary.

E-mail Policy

I have a simple email policy, and it is targeted at achieving one goal: **maximizing your and my work-life balance**. The policy is this:

I promise a timely response to **relevant** emails I receive between **9:00 AM** and **5:00 PM** Monday to Friday.

You may not think professors have lives (but we in fact do, and I like to live mine outside of normal

working hours). That doesn't mean I expect students to abide by chrono-normative standards.[2] But, this does mean that if you email me outside of these windows, I may not respond until the next 9-5 workday.

Make-Up Assigment Policy

There are **NO** make-ups for missed assignments. Don't bother asking. But, if you anticipate having troubles making a due-date and notify me ***in advance***, we can work out a solution.

Attendance Policy

Come to class. Attendance is not part of your grade, but poor attendance will lead to poor grades.

Computer-based Excuses

Excuses for late or missed assignments based on CD, flash drive, or hard drive errors are **not acceptable**. The Denison network and server is reliable and accessible. If you use your Google Drive or some other cloud-based platform to save your work, all your work will be backed up and easy to access.

Late Assigments

We have a lot of work to do in this class. So turn in your work when it's due. This is meant to help you. I love to procrastinate just as much as anyone else—but if you procrastinate in this course, you will drown. As incentive for keeping on top of your assignments, each day your assignment is late and unexcused (including weekends and holidays) you will lose 5 percent from your final grade for that assignment. There will be no exceptions made for work that is submitted only minutes after a deadline. If something is due by midnight on a Friday (12:00 AM), the moment the clock strikes 12:01 AM your assignment is a day late and you will automatically lose 5% from your grade.

The exception to this rule is if a student and I have worked out an arrangement for submitting an assignment at a later date.

Electronic Submission

You will submit all of your assignments electronically via Canvas.

Academic Dishonesty Policy

Don't cheat. Just don't do it.

It should go without saying, but ***plagiarism*** is also a form of cheating and it includes:

1. Copying or paraphrasing the ideas of others without citation or attribution.
2. Copying or paraphrasing the ideas of ***other students in the class***.

[2] By this, I simply mean classic societal expectations about working vs. leisure hours.

I've had to deal with students plagiarizing before. It's painful for me and it puts a blight on the record of the student. It's not only cheating, it's stealing.

When in doubt about whether something constitutes cheating, consult Denison's Code of Academic Integrity.[3] Be advised that this same Code of Academic Integrity requires that instructors notify the Associate Provost of cases of academic dishonesty. **Any incidence of academic dishonesty will result in failure of the course and referral to the Denison judicial process.**

[3] Of course, if you have to ask yourself if something counts as cheating, then it probably is...