

## Introduction to Data Science - Fall 2020

### Class and Instructor Details

STAT 111-004 (3 credits)

Virgil U. Pierce

Email: [virgil.pierce@unco.edu](mailto:virgil.pierce@unco.edu)

Cell: 956-249-0566

Office: Ross Hall 2239

Class Meetings: Monday, Wednesday, Friday 10:10a - 11:00a Via Zoom.

The Zoom Link is:

<https://unco.zoom.us/j/97351253002?pwd=aTlMTndUUjFHMzUxYjJSRitwQlIQUT09>

The password will be sent in an email. The direct link and dial in options will be in Canvas.

We will discuss the specifics of our schedule on Monday 8/24 in our first meeting; but the plan is to begin with 2-3 days per week of synchronous meetings, transition to 1 synchronous meeting per week by the end of the semester. In plain language: keep the class meeting time free in your schedule for now, but anticipate that as the semester goes on you will have more flexibility and freedom. Because there are only four of us, we can also arrange to meet either one-on-one or as a class if that is what we decide to do.

Note that for Fall 2020 our enrollment is small enough that we will make some decisions during the course about what we want to do as a group. The plan for the course is to have a series of projects that you complete in between exploring the tools we will use. We will need to decide between us whether we do those projects together or whether we do them individually.

Youtube Playlist: Videos for class, including recordings of class will be put into a youtube playlist available here.

<https://www.youtube.com/playlist?list=PLtO14BSFFYs7MreUbDcGgqDAbwycH3RPw>

Student Hours: TBD. You can always email me and we can schedule a meeting. Emails will have a response within 24 hours. *See the note below about this democratic syllabus.*

### No Textbook Purchase Required.

### LAC Area Met by the Course

The course meets all 10 objectives for Area 2 of the UNC Liberal Arts Core.

## **Course Description and Prerequisites**

From the catalog: Introduction to elements of Data Science and elementary tools, programming languages, and techniques for data collection, visualization, computations, and inference. Includes ethical issues with data collection and analysis.

A more complete description: This course is an introduction to the elements of Data Science with an emphasis on the elementary tools and techniques needed for data collection, visualization, computations and inference. Students will gain hands-on experience working with data using modern computational tools and programming languages with direct applications in future data science and statistics courses and careers. Students will consider the ethical issues with data collection and analysis.

Pre-requisites: A score of 25 or higher on the ALEKS Placement Assessment, successful completion of an entry level college mathematics courses, or permission of the instructor.

## **Textbook and Software**

Both of our texts and the software we use for class will be available online. The reference text is available as a PDF which you could print if you needed to. If you have accessibility concerns about the texts, please let me know, in particular the primary text should play nicely with a screen reader and the reference text is something we can have translated into a screen reader friendly format.

Primary Text: Computational and Inferential Thinking, by Ani Adhikari and John DeNero.  
<https://www.inferentialthinking.com/chapters/intro.html>

Reference Text: Introductory Statistics with Randomization and Simulation, by David M. Diez, Christopher D. Barr, and Mine Cetinkaya-Rundel.  
[https://www.openintro.org/download.php?file=isrs1\\_tablet](https://www.openintro.org/download.php?file=isrs1_tablet)

The course has a programming component, but let me lead with: we will take it very gently. This course is being designed as an appropriate class for someone who has not taken a programming class before. It will serve as an introduction to programming by having you focus on tasks associated with data collection and analysis.

The softwares we will be using are the Python programming language and Jupyter Notebooks (an interactive environment for Python). You can install Python and Jupyter on your own computer (instructions in Canvas), however, the university has a server that is running a Jupyter Notebook with Python that will be available for you. This will let you use Python without having to install it, and you could also access it using a tablet or phone.

You can access the Jupyter Hub by going to this website:

<https://jupyter.unco.edu/hub/login> click sign on with Azure ID (you may be asked to login with your UNC credentials at this step) and then choose the minimalist Python notebook (this keeps our resources usage lower). That's it. Your account should be persistent, unfortunately sharing files is not as easy as google - you will need to download them to your own machine. I will distribute instructions for getting the files we will use during class.

Jupyter Notebooks are becoming a common tool for people doing data science and allow the creation of high quality documents with the code and results of computations directly embedded (for instance our primary text). Python is quickly becoming the standard language for working with data, particularly when the amount of data becomes large enough that traditional methods like spreadsheets stop working.

The introductory videos will include some information about why Python is becoming the dominant tool and what alternative software is available.

#### Outline of Course Content:

<b>Week 1</b>	Motivation - Case Study 1
<b>Week 2</b>	Using Python and Jupyter - Importing and Working with Data
<b>Week 3</b>	Data Types and Summaries
<b>Week 4</b>	Data Visualizations
<b>Week 5</b>	Communicating Results - Case Study 2
<b>Week 6</b>	Data Engineering - Writing Functions, Loops, Controls
<b>Week 7</b>	Randomness and Sampling
<b>Week 8</b>	Testing Hypothesis
<b>Week 9</b>	Comparing Two Samples - Case Study 3
<b>Week 10</b>	Estimation and Confidence
<b>Week 11</b>	Making Predictions
<b>Week 12</b>	Regression
<b>Week 13</b>	Classification
<b>Week 14</b>	Making Predictions - Case Study 4

<b>Week 15</b>	<b>Presentations of Final Case Studies</b>
----------------	--

### **Learning Objectives/Outcomes for the Course**

As a result of this course, students will expand their skill set to work and think with data.

They will:

1. Demonstrate good problem-solving habits with novel datasets: exploring the dataset with visualization and summary statistics, experiment with the tools from class to determine the best approach, and interpret the results of their analysis;
2. Generate and interpret a dataset and their analysis of it using symbolic, graphical, numerical, and verbal representations of the mathematical statistical ideas;
3. Communicate results of their analysis, including explaining the data science workflow and why their analysis was structured the way it was;
4. Apply machine learning methods and assess the quality of predictions;
5. Recognize and apply the appropriate methods and analysis to datasets;
6. Manage, summarize, and visualize data using R or Python;
7. Demonstrate competency in Quantitative Literacy by being able to:
  - a. Interpret summaries and visualizations of a dataset,
  - b. Represent information from a dataset in a variety of ways,
  - c. Perform statistical inference and implement simulation-based inference methods on a dataset,
  - d. Communicate the results of their analysis, and
  - e. Address the assumptions underlying their analysis; and
8. Evaluate and describe the ethical concerns with a data collection and analysis project.

### **Communication**

We will be using Canvas to communicate for this class. Particularly the discussion boards are a useful place to post questions and collaborate with other students. Unless your question for me is about your grades, I would love to have you post it in the discussions page in Canvas. I can guarantee if you have a question, someone else has the same question.

There are only four of us, so we can also decide to use some other method to communicate. Email is problematic for me as the volume I receive is more than 100 messages per day, and it is more effective for my teaching if student communication is in another channel.

### **Democracy**

My classroom functions a little bit like a democracy. One example is that in the first week we will decide when the best hours during the week for student hours are (my partner has veto authority here).

At any time you can bring up a proposal for the class, and it can be discussed and voted on. With only four of us, mutual agreement is necessary for all changes. Not all proposals will be things I can act on for one reason or another, but I am open to discussion about any decision related to your class. Proposals that are accepted become acts and will be added to this syllabus.

Some examples:

1. I have had classes change the due dates on assignments;
2. Change the weights for assignments;
3. Adjust the content for an exam.

### **Homework**

Homework is a fundamental part of this class. You are free to work with others on the class homework (and with only four of us, you should), but all work submitted must be your own, and you may be asked to explain your reasoning, or the method used on a particular problem. There will roughly be one or two homework assignments per week, and I would budget 3-4 hours per week for them. Note that you do not necessarily need to have a perfect solution - for one thing there rarely is a perfect solution for this class, but rather you should have spent some time thinking about how to get started.

### **Case Studies**

There will be four case studies done as part of class. These are where you will combine the techniques we have been learning in class to a novel dataset (well except for Case Study 1 which is our motivating example). The emphasis is on using the tools we are learning about, exploring issues that come up, and communicating our results. You will submit a written report of each case study.

### **Class Videos and Presentations**

You will be asked to show the class what you are working on and how you are thinking about the class activities and homework. Occasionally these will be recorded videos you make outside of class. I know that making videos takes time, so I will assign these infrequently.

### **Projects and Final**

There will be three projects and a final that will build on your progress in the course. They will be less involved than the case studies and will instead ask you to reflect on a data analysis result and evaluate the project. The final will involve expanding on your work in one of the case studies and then a reflection about the class and what you have accomplished. I am calling these projects because they are being designed for the online delivery of this class and will not be tests or exams in the traditional sense. You will have five days to complete these assignments and a week to complete the final project.

## **Grading Policies**

Your grade in the class will be based on homework, the case studies, class presentations, and tests and a final. The final grade will be a weighted average using:

- Homework: 30%
- Case Studies: 20%
- Class Presentations: 20%
- Projects and Final: 30% (3 projects and a final will have equal weight)

Final Grades will be based on a percentage of this weighted total. The distribution will be no tougher than: A (90- 100%); B (80- 89%); C (70 - 79%); D (60 - 69%); and F (1 - 59%).

## **UNCO Policy Statements**

### **Disability Support Services**

It is the policy and practice of the University of Northern Colorado to create inclusive learning environments. If there are aspects of the instruction or design of this course that present barriers to your inclusion or to an accurate assessment of your achievement such as time-limited exams, inaccessible web content, or use of videos without caption, please communicate this with your professor and contact Disability Support Services (DSS) at (970) 351-2289, Michener Library L-80 to request accommodations. Students can learn more about the accommodation process at

<http://www.unco.edu/disability-support-services/accommodations/>

### **Honor Code**

All members of the University of Northern Colorado community are entrusted with the responsibility to uphold and promote five fundamental values: Honesty, Trust, Respect, Fairness, and Responsibility. These core elements foster an atmosphere, inside and outside of the classroom, which serves as a foundation and guides the UNC community's academic, professional, and personal growth. Endorsement of these core elements by students, faculty, staff, administration, and trustees strengthens the integrity and value of our academic climate.

### **UNC's Policies**

*UNC's policies and recommendations for academic misconduct will be followed. For additional information, please see the Student Code of Conduct at the Dean of Student's website <http://www.unco.edu/dos/Conduct/codeofconduct.html>. In the case of academic appeals, university procedures will be followed. For information on academic appeals, see <http://www.unco.edu/regrec/Current%20Students/AcademicAppeals.html>.*