BUSN 5000
Introduction to Data Science for Business and Economics

Chris Cornwell

Spring 2024
(updated 20 Dec 23)

Class Room: Amos B010
Recitation Room: Ivester 007

Class Hours: TR, 935a and 1110a
Recitation Hours: R, 545p-645p

Teaching Team

Instructor

Contact	Chris Cornwell
web	https://cornwl.github.io/
email	cornwl@uga.edu
office	Amos B414
hours	W, 200-330

Teaching Assistants

Contact	Abbi Cormier	Morgan Kearns
email	abigail.cormier@uga.edu	morgan.kearns@uga.edu
office	Orkin 207	Amos B413
hours	T, 200-330	by appt

Abbi is the lead TA and the primary contact for questions about the course administration and assessments. Morgan is the recitation leader and will be conducting regular sessions to give students an opportunity outside of class to revisit the course material and ask questions.

Course Description

The modern world is awash in a seemingly unlimited amount of data. To harness these data for decisions starts with acquiring the raw information and ends with a report describing the result of some analysis. At each stage of the value chain, the analyst combines data with ideas about how the world works to produce an output. BUSN 5000 will take a hands-on approach, covering data preparation; descriptive, explanatory and predictive analyses; and scientific communication.

Course Objectives

After completing this course, you should understand how to

1. acquire and prepare data for analysis.
2. design reproducible data analyses.
3. map business problems and policy questions to hypotheses about relationships in data.
4. describe data and perform basic descriptive analysis.
5. implement and interpret basic causal-inference research designs.
6. implement and interpret basic machine-learning algorithms.
7. communicate the results from descriptive, causal and predictive analyses.

Topical Outline

The topical outlines for parts I and II of the course are provided below. A detailed course schedule is posted on eLC and here.

Part I :: Transformation to Analysis

1. Data fundamentals
2. Beginning to learn
3. Models for exploration
4. Making inferences
5. Measurement error, sample selection, and confounding
6. Bayesian approach to learning from data
7. Data ethics

Part II :: Explaining and Predicting

1. Regression fundamentals
2. Potential outcomes and causal inference
3. Regression discontinuity
4. Difference in differences
5. Prediction with regression
6. Introduction to machine learning

Recommended Texts

There are no required texts for this course, but many useful ones. Here is a curated list where you can find the course content covered at an "age-appropriate" level. We highly recommend the "Gabors" book for this class.

- *Beginner*

  Bueno de Mesquita, E. and Fowler, A., *Thinking Clearly with Data*, Princeton University Press.

  Çetinkaya-Rundel, R., *Data Science in Box*.

  Healy, K., *Data Visualization: A Practical Introduction*, Princeton University Press.

- *Intermediate*

  Bekes, G. and Kedzi, G., *Data Analysis for Business, Economics, and Policy*, Cambridge University Press.

  Angrist, J. and Pischke, S., *Mastering 'Metrics*, Princeton University Press.

- *Next Level*

  Cunningham, S., *Causal Inference: The Mixtape*, Yale University Press.

  Schwabish, J., *Better Data Visualizations*, Columbia University Press.

If you need a statistics refresher, we strongly recommend *Computational and Inferential Thinking: The Foundations of Data Science*. It is the foundational text for Berkeley's Data 8 course, which is a modern take on the topics covered in BUSN 3000. Note that the coding component of the course is based in Python.

There are many other good online resources for statistics coverage with R programming tossed in. Check out *R for Data Science*, *Learning Statistics with R*, and *Foundations of Statistics with R*.

Software

The software of choice for this class is R, a free and open-source language for statistical computing and graphics. RStudio is a popular integrated development environment (IDE) for R that will greatly enhance your R experience. First, download and install R; then download and install RStudio. Follow these instructions.

Terry Analytics Lab

The TAL is a resource supported by the Center for Business Analytics and Insights for all Terry students enrolled in business analytics and econometrics courses. TAL is managed by *Taylor Gorak* (taylor.gorak@uga.edu). Her team provides free tutoring in course concepts and R coding. TAL typically operates from early/mid afternoon to early evening, Mon-Thu, beginning the second week of the semester. You should regard TAL as the first stop for help with course content and assignments. TAL has its own eLC course page to which you will be subscribed.

Course Policies

Performance evaluation

Your performance will be evaluated on the basis of homework assignments, Dailies, projects and a final exam weighted as follows:

Assessment	Number	Total
Dailies		10%
Homework	10	30%
Project	1	20%
Test	2	40%

Dailies

A Daily comprises 1–3 short active-learning exercises that we will conduct each class period. You will earn Dailies points by participating and providing correct responses. Your overall Dailies score will be the percentage of total potential Dailies points you earned. *If you are absent from class you cannot participate in the Daily.* (This should go without saying, but just in case: participating in Dailies on behalf of students who are absent from class constitutes an academic honesty violation and will be prosecuted as such.) We will drop the 4 lowest Dailies scores.

Homework

Homework assignments are *formative* graded tutorials that guide you through the key concepts in each course topic and include an empirical component involving R. We will drop the 2 lowest homework scores.

Homework assignments are delivered as a Shiny app running in the cloud. We have posted a video explaining the process of submission on eLC. Watch it and then watch it again. In addition, you must complete *Homework0*, which gives you the experience of completing and submitting a Homework assignment. Homework0 does not contribute to your course grade, but we will not acknowledge any of your Dailies or real Homework submissions until you complete it.

Project

The Project is a *summative* assignment that draws on key course concepts to learn about an empirical relationship and document what you learn in a notebook-style deliverable. The Project is to be completed individually. You should not collaborate with other students, but you may consult TAL staff or the TAs for assistance with the coding elements and knitting your R Markdown document into the required notebook format.

We have also posted a video explaining the process of submitting the project on eLC. You should probably watch this one at least 3x. In addition, you must complete a *pre-Project exercise* that takes you through the steps of establishing a workflow, knitting an R Markdown document, and creating a PDF version for submission. This exercise does not contribute to your course grade, but we will not acknowledge any of your Dailies or Homework submissions until you complete it.

Tests

The tests are *summative* assessments of the key concepts covered in each section of the course. We believe real mastery of the course material is demonstrated by high performance on both formative and summative assessments.

The tests will be held in class, the first on **Feb 29** and the second during the final exam period according the each section's prescribed date and time. The 930a section will take the second test on **May 2 at 800a** and the 1110a section will take the test on **May 7 at 1200p**. You must take the second test on the date for your section.

Dates, deadlines and drops

If you know now that you will not be able to take either test on the date it is scheduled, you should drop this class.

We are likewise serious about Homework and Project deadlines, which are indicated in the course schedule. *Late submissions will not be accepted.*

The policy of dropping the 4 lowest Daily and 2 lowest Homework scores is to accommodate unforeseen events that prevent you from participating in a particular Daily or submitting a particular Homework assignment. Think of the policy as insurance. The drops are your allotted claims. Use them wisely because when they're gone, they're gone.

Attendance

There is no explicit penalty assessed for missing class, but missing class means missing a Daily. Make no mistake, the data on Dailies participation show that regular class attendance predicts success in BUSN 5000.

Overall assessment

You will be ranked relative to other students in the class according to your overall performance and grades will be assigned based on your class rank. We will use the plus/minus system to make distinctions within grade categories.

Communication

Our communications to the class will generally come through the eLC Announcements tool, which functions like an instant messaging system. You should set your notifications preferences to receive Announcements postings in the manner that suits you. We strongly encourage the SMS option. Regardless, you are responsible for information conveyed in the announcements.

Please address content and assignment-related questions to TAL staff first. If TAL staff members cannot solve the problem, please reach out to the TAs or to me.

Course administration questions should first be directed to Abbi. If she is unable to resolve the issue, I will be happy to intervene.

When you write to any member of the teaching team, your email *must* have the following components:

- A subject line that includes your section time (935a or 1110a) and an few words that categorize the problem (e.g. "coding error" or "homework question")
- A proper greeting ("Hey," is not a proper greeting. "Dear Abbi" or "Dear Ms Cormier" is.)
- A clear description of the problem. If it involves a coding issue, include your code (the relevant chunk from a Homework assignment or .Rmd file for the Project). Do *not* send a screenshot or photo of the problem.
- A proper closing (e.g. "Respectfully, your_name")

Omission of any of these features may cause your message to be rejected.

Electronic devices

Cell phones *must* be muted or turned off and stowed away during class. Laptops may be used in class, but only for purposes directly related to the course (e.g., taking notes, live coding and viewing course materials).

Generative AI

Our view is that generative AI (ChatGPT, Claude, Bard, etc) is productivity-enhancing tool and you should learn to use it in this course and in life. One way to do this is integrate GitHub Co-pilot into RStudio. Here is a video that explains how to use it. There is a $10/month or $100/year fee. If you are interested, you can get a free trial here.

University and College Policies and Statements

UGA student Honor Code

"I will be academically honest in all of my academic work and will not tolerate academic dishonesty of others." A Culture of Honesty, the University's policy and procedures for handling cases of suspected dishonesty, can be found at https://www.uga.edu/ovpi.

Mental health and wellness resources

If you or someone you know needs assistance, you are encouraged to contact Student Care and Outreach in the Division of Student Affairs at 706-542-7774 or visit https://sco.uga.edu. They will help you navigate any difficult circumstances you may be facing by connecting you with the appropriate resources or services.

UGA has several resources for a student seeking mental health services (https://www.uhs.uga.edu/bewelluga/bewelluga) or crisis support (https://www.uhs.uga.edu/info/emergencies).

If you need help managing stress anxiety, relationships, etc., please visit BeWellUGA (https://www.uhs.uga.edu/bewelluga/bewelluga) for a list of FREE workshops, classes, mentoring, and health coaching led by licensed clinicians and health educators in the University Health Center.

Additional resources can be accessed through the UGA App.

Diversity and inclusive excellence

The Terry College of Business is committed to promoting an inclusive learning and working environment among its students, faculty, and staff. This class welcomes the open exchange of ideas and values freedom of thought and expression and provides a professional environment that recognizes the inherent worth of every person. It aims to foster dignity, understanding, and mutual respect among all individuals in the class.

Changes to the syllabus

The course syllabus is a general plan for the course; deviations announced to the class by the instructor may be necessary.