BUSN 5000E
Introduction to Data Science for Business and Economics

Chris Cornwell

Summer Thru Term
(updated 16 May 25)

Class Room: Online | Class Hours: Asynchronous

Welcome to BUSN 5000E

BUSN 5000E is the online, asynchronous version of BUSN 5000 and is every bit as rigorous. The description, learning objectives, and topical outline are the *same*. The assignments have been adapted to the online format and the shorter calendar, but they are designed to assess the *same thing*. The message here is that you should not confuse the "E" with "easier". If anything, the challenge will be greater because you will not have the disciplining forces associated with regular in-person class meetings and recitations.

Your success in BUSN 5000E will depend heavily on your motivation, commitment and organization. You should start by doing 2 things asap:

1. Download and install R and R Studio (directions below).
2. Carefully review the course schedule and incorporate it into your summer calendar. Don't gloss over the deadlines policy. We are serious about that.

This course follows the schedule of the Summer Thru Term. All required course materials are available through eLC.

Teaching Team

My TA and I will be available to guide you and address your questions as they arise. You should see us as facilitators, tutors and coaches. In addition, we've built a GPT that is trained on the course content and will answer any question you put to "him". We call him "Don". Don is not flawless, but he's available 24/7 at the link below.

Contact	Web	Email	Hours
Chris Cornwell	https://cornwl.github.io/	cornwl@uga.edu	TBA
Abbi Cormier (TA)	https://abbicormier.github.io/	abigail.cormier@uga.edu	TBA
Don	BUSN 5000 GPT 2.0		24/7

Abbi is the primary contact for questions about course grades and administration.

Course Description

The modern world is awash in a seemingly unlimited amount of data. Harnessing these data for decision-making begins with acquiring the raw information and ends with communicating the results of analysis. Along the way, the data are transformed for analysis and the analyst matches statistical methods to the task at hand. BUSN 5000 covers the data science skills necessary at every stage of the value chain, including data transformation; descriptive, explanatory and predictive analyses; and professional communication.

Student Learning 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.

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

Part II :: Explaining and Predicting

7. Regression fundamentals
8. Potential outcomes and causal inference
9. Regression discontinuity
10. Difference in differences
11. Prediction with regression
12. Introduction to machine learning

A detailed course schedule is posted on eLC and here. You should plan to follow it to a "T". Each topic of the twelve topics corresponds to a module on the schedule.

Course Materials

Course notes packet

The BUSN 5000E slide decks are written to be read rather than presented. You should regard them more like a textbook. To make them more useful, we have arranged with Bel-Jean's to offer

printed copies in a notes format (3 slides per page with space for note-taking). The packets are available in black and white (cheaper) or color (more expensive). The purpose of the packet is to provide structure for taking notes as you study course content. We strongly recommend that you purchase a copy use them in this way. Viewing them on your laptop would be beside the point.

Videos

Most of the technical content in the slide deck packet is supported by instructional videos. They are generally 5-10 minutes long and designed to capture the steps I would take in class, working at the whiteboard, to explain a particular concept. Each is mapped to a corresponding slide or slides by its title.

Recommended texts

There are no required texts for this course (other than the slides packet), but many useful ones. Here is a curated list where you can find the course content covered at a "skill-appropriate" level. If we were to recommend just one to buy, it would probably be the "Gabors" text, as it has the most comprehensive treatment of the topics covered in BUSN 5000E. Anyone considering continuing on from BUSN 5000E to Terry's MSBA or a similar program should buy it. †The schedule shows the relevant sections from the Gabors text, as well as those from Bueno de Mesquita and Fowler and Angrist and Pischke.

- *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, you might start with *Computational and Inferential Thinking: The Foundations of Data Science*. It is the text for Berkeley's Data 8 course, which is a modern take on the topics covered in introductory statistics courses. It includes basic instruction in Python as well.

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*. Finally, there is *R for Excel Users* for those wanting some guidance in transitioning from spreadsheets to a proper scripting language.

IRLs

In addition to the text recommendations, each module is associated with some IRL ("in real life") content – articles, podcasts, and videos – that ties the course material to real-world settings. With the exception of the Uber story in module 2, there is no direct accountability for the IRLs. They are offered for your edification.

Software

Data analysis in this class is done in 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. Our *helpR* guide posted on eLC links to numerous resources for learning R.

Course Policies

Performance evaluation

Your performance will be evaluated on the basis of homework assignments, Dailies, projects, and in-class tests weighted as follows:

Assessment	Number	Total
Homework	10	30%
Project	1	20%
Exam	1	50%

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 Shiny apps running in the cloud. Links to each assignment are posted on eLC. We provide access to Shiny's cloud service at no cost to you and you do not need an account to access it.

We have posted a video explaining the process of submission on eLC. Watch it and then watch it again. Submissions that do not comply with the process will not be accepted. To promote compliance, we require you to complete *"Homework 0"*, which gives you the experience of completing and submitting a Homework assignment. Homework 0 does not contribute directly to your course grade, but failure to complete it by the deadline will eliminate one of your two drops.

Project

The Project is a *summative* assignment in which you draw on key course concepts to learn about an empirical relationship and document what you learn. You will use R and R Markdown to conduct the analysis and report your findings, "knitting" the two together in a slide deck. The Project is to be completed individually. You should not collaborate with other students, but you may consult the TA or me for coding and knitting assistance.

We have posted a video explaining the process of submitting the project on eLC. You should probably watch this one at least 3x. Submissions that do not comply with process and the rubric spelled out in the project instructions will not be accepted. To promote compliance and provide the foundation for a successful project, we require you to complete a *Pre-Project exercise* that takes you through the steps of establishing a workflow, knitting an R Markdown slide deck, and creating a PDF version for submission. *This exercise accounts for 10% of your Project grade and must be completed by the deadline indicated on the course schedule*.

In addition to the Pre-Project exercise, we offer an optional *Project Progress Check* in Week 4 that covers about two-thirds of the project tables and figures. We will award 5 bonus points to your project grade if your project progress check submission is complete and correct.

Exam

The exam is *summative* assessment of the key concepts covered in each module of the course. We believe real mastery of the course material is demonstrated by high-level performance on both formative and summative assessments.

The exam will be accessible from 600a–1159p on Fri, Aug 1 (all times are EDT). Once you begin the exam, you will have 150 minutes to complete and submit it. See the exam guide posted on eLC for details of administration and instructions for completing and submitting it.

The exam accounts for half (50%) of your total grade. Why such a high-stakes assessment? I believe real mastery of the course material is indicated by a strong performance on a timed, summative assessment like a cumulative final exam. Placing a large weight on the final is intended to promote effort toward preparing for that assessment. The homework assignments are low-stakes formative assessments designed to prepare you for that final demonstration of mastery. Because the homework assignments are untimed and open-book, they provide opportunities to earn good scores. In a similar fashion, the project is designed for you to successfully complete it and earn a good score. Together, the homework assignments and project comprise the other half of your course grade, balancing out the high stakes of the final.

Dates, deadlines and drops

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

The policy of dropping the 2 lowest Homework scores should accommodate unforeseen events that prevent you from submitting a particular Homework assignment by its deadline. We don't distinguish between excused and unexcused drops, so there is no reason to email anyone on the teaching team to document your reason for dropping an assignment, whether it be for illness, an interview, car trouble, or whatever. Just take the drop and move on. Think of the policy as insurance. The drops are your allotted claims. Use them wisely because *when they're gone, they're gone*.

We are likewise serious about project deadlines and exam dates. *Late pre-project exercise and final project submissions will not be accepted* and *all students are required to take the exam during the prescribed window*. If you know now that you will not be able to take the exam during the period it is scheduled, you should drop this class.

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. We do not "round up".

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. Regardless, you are responsible for information conveyed in the announcements.

Address content and assignment-related questions to the TA or me.

Direct course administration questions first to the TA. 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 and a 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 and how you have tried to solve it. If it involves a coding issue, include the code in your message. For homework assignments, paste the relevant chunk in the body of your email. For the project, attach your `.Rmd` file and paste the error message in the body of your email. Do *not* send screenshots or photos.
- A proper closing (e.g. "Respectfully, your_name")

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

Electronic devices

The in-person version of BUSN 5000 has a strong *no-electronic-device policy*. Why am I mentioning this here? Because we're convinced that phones are a clear distraction to learning. The data (here, here, and here) clearly indicate their use in class harms learning and learning is what we care about. We pretty confident this is true outside of class as well, so we strongly recommend you put your phone away when working on this course.

Generative AI

My take is that you should view generative AI chatbots, like ChatGPT, Claude, and Gemini, as exceedingly productive collaborators with whom you should learn to work, much like you would with a human. Unfortunately, there are no secret prompting strategies or special step-by-step manuals (also like with a human). For some practical guidance along these lines, I recommend that you follow Ethan Mollick's substack.

As you probably have discovered, one place gen AI really helps is in coding. You can get coding help from the chat interface but you may prefer using Cursor or integrating GitHub Co-pilot with your favorite IDE. Here is how to get free access to Github Co-pilot as a student. Here are instructions for integration into RStudio. Alternatively, you may find you can get by just vibe coding with the chatbot.

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.

UGA Well-being Resources

UGA Well-being Resources promote student success by cultivating a culture that supports a more active, healthy, and engaged student community.

Anyone needing assistance is encouraged to contact Student Care & Outreach (SCO) in the Division of Student Affairs at 706-542-8479 or visit https://sco.uga.edu. Student Care & Outreach helps students navigate difficult circumstances by connecting them with the most appropriate resources or services. They also administer the Embark@UGA program which supports students experiencing, or who have experienced, homelessness, foster care, or housing insecurity.

UGA provides both clinical and non-clinical options to support student well-being and mental health, any time, any place. Whether on campus, or studying from home or abroad, UGA Well-being Resources are here to help:

- Well-being Resources: https://well-being.uga.edu
- Student Care and Outreach: https://sco.uga.edu
- University Health Center: https://healthcenter.uga.edu
- Counseling and Psychiatric Services: https://caps.uga.edu or CAPS 24/7 crisis support at 706-542-2273
- Health Promotion/ Fontaine Center: https://healthpromotion.uga.edu
- Disability Resource Center and Testing Services: https://drc.uga.edu

Additional information, including free digital well-being resources, can be accessed through the UGA app or by visiting https://well-being.uga.edu.

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.