**PDAT 610G Introduction to Data Science**

**Course Syllabus**

Last Revised: 8/12/21

**Instructor:** Dr. Hyun-Joo Kim  **Academic Success Mentor:** Andrea Maag

**Email:** hjkim@truman.edu **Email:** amaag@truman.edu

**Availability:** R 6:30-7:30 pm, **Phone**:660-785-7403

WF 8:30-10:00 am, **Availability:** Byappointment

Or by appointment

**Welcome to Truman’s Data Science Graduate Program**

I am excited to have you in Truman’s data science program and PDAT 610G. This is the first class of the program, introduction to data science. “Look around you” (this might actually happen in a zoom setting) and these people are the ones that you will work together throughout. In this class, in addition to an introduction to R and overview of common statistical analysis, we will learn about each other and the program, and work on your degree plan. I am your PDAT 610G instructor, academic advisor, and a helper for your success during your graduate program. As you start the new certificate/master program, please let me know if you have any question.

**Zoom meeting and office hours**

Our weekly zoom meeting will be on Thursday 6:30 pm -7:30 pm. Although it is not mandatory, I strongly encourage you join us and participate the conversation in the meeting. I might record some part of the meeting, but you will really get the full experience by being there.

On August 19th**,** our very first meeting, you will meet the instructor (me), Andrea (your academic success mentor) and other students in the class. It is often really helpful to know others to be successful in the class. This will be a great chance to meet everybody especially since you will take many classes together as a cohort.

I will be available via zoom on Thursday 6:30-7:30 pm and WF 8:30-10:00 am. You are welcome to email me with questions anytime and I will answer you as soon as I can. I will be also available via zoom during office hours if you want to ask me a question in person. Just shoot me an email when you want to talk to me.

**TECHNOLOGY REQUIREMENTS**

In order to participate the course fully and effectively, students should have a reliable broadband connection (Cable Modem, DSL, Satellite).  Note that Windows is ending technical support and security updates for Windows 7.  Students should have a relatively new operating system (Windows 7 or newer; Mac OSX, etc.) and employ a compatible browser such as Chrome or Safari.   Courses use Blackboard Learn.  For a list of compatible systems and browser types, visit Blackboard.

**Description**

This course introduces students to the world of data science by providing an overview of the discipline, in-depth exposure to the statistical software environment R, and an overview of common statistical analyses. Upon completion of this course, students will have a solid foundation upon which to build in completing the remainder of the courses that make up the data science certificate program.

**Prerequisites**

Successful completion of an introductory computer science course, such as CS170 Introduction to Computer Science I or CS180 Foundations of Computer Science I, and STAT190 Basic Statistics, or their equivalent in coursework or experience.

**TEXTBOOK**

There is no required textbook for this course. Students may find *The* *Book* *of* *R:* *A* *First* *Course* *in* *Programming* *and* *Statistics*, Tilman M. Davies, No Starch Press, San Francisco, 2016 useful as a reference.

**COURSE OBJECTIVES**

By the end of this course, the successful student will:

* Achieve a basic level of knowledge regarding the data science discipline
* Develop an understanding of statistical reasoning used in the field of data science
* Demonstrate proficiency using the R programming language using the front‐end interface RStudio
* Demonstrate the ability to recognize and create quality visuals that accurately present research results

**COURSE CONTENT**

1. **Overview of Data Science and PDAT**

* What is data science and why is it important?
* What do data scientists do? Why are they in demand?
* Getting Started
  + Overview and history of R
  + RStudio console and various windows
  + Libraries and R Resources
  + R markdown

1. **The base R Language and Tidyverse**
   1. **How R thinks about data**

* Installing R and RStudio
* Data Types
* Subsetting
* Vectorized operations
* Importing and Saving Data.
  1. **Welcome to the Tidyverse**
* Literate and Reproducible Programming
* What is the Tidyverse?
* Data Cleaning, with dplyr: filter, select, arrange, mutate, summarize, with group\_by

1. **Programming**

* Using an existing function and creating a new function
* Conditions and loops -- if else, for, while
* Loop functions – apply, lappy, sapply, mapply
* Project part I

1. **Statistics and Probability**

* Elementary statistics
  + Measure of tendency: mean, median, etc.
  + Frequency distributions and cumulative frequency distributions
  + Histograms and Scatter plots
* Probability
* Common probability distributions: Binomial trials, Normal distribution, Uniform distribution

1. **Overview of Statistical Testing and Modeling**

* Sampling distributions and confidence/central limit theorem
* Hypothesis testing
  + One-sided and two-sided tests for one sample
  + One-sided and two-sided tests for two samples
* Simple linear regression and diagnostics

1. **Advanced Graphics**
   * Base R
   * Grammar of Graphics (ggplot)
   * Advanced plot customization
   * Going further with the grammar of graphics
2. **Project part II**

**GRADING**

There are 8 modules for 8 weeks and typically later modules are longer than earlier module. Finishing one module a week may be the best for you to successfully digest the material, learning assessments include practical application programming assignments and discussion for each course module. In particular, completing the discussion in a reasonable time frame is important for you and the entire class to have more effective and productive conversation. Please follow the due dates.

|  |  |
| --- | --- |
| Modules | Due dates |
| Module 1 | Aug. 22nd |
| Module 2a | Aug. 29th |
| Module 2b | Sept. 5th |
| Module 3 | Sept. 12th |
| Module 4 | Sept. 19th |
| Module 5 | Sept. 26th |
| Module 6 | Oct. 3rd |
| Module 7 (Project) | Oct. 12th |

Practical application programming assignments will allow you to practice the skills of the course. Programs will be graded on adherence to specifications, program quality, and correctness of interpreting results. Quizzes may be provided during the lecture to monitor your own progress throughout the course.

In the course discussions, you are asked to post your own answers to discussion questions and to read the posts of other learners and provide them with feedback. Thus, you will contribute for each discussion: an initial post that addresses the discussion question and a feedback post to one or more of your peers' initial posts.

Your overall grade will be defined:

|  |  |
| --- | --- |
| Assignment 100 points \* 7 | 700 |
| Discussion 50 points \* 8 | 400 |
| Project | 100 |
| Total | 1200 |

**90% and up -------------------A**

**80% and up -------------------B**

**70% and up -------------------C**

**60% and up -------------------D**

**59% and down ----------------F**

An overall course grade lower than a "C" is considered a failing grade, and no more than one course grade of "C" may be counted toward the certificate. Students may retake a course to raise a grade not meeting minimum program requirements.

**My Expectations of Students**

I expect students to: thoroughly read the required materials (watch the lecture video) and complete assigned activities prior to engaging in discussion boards; complete their assigned work; engage in active conversation with peers; engage in polite and respectful discourse with their peers; ask questions whenever they have difficulty understanding material or course requirements; express concerns about the course at any time.

**What Students Should Expect of Me as Their Instructor**

Students should expect me to: provide timely feedback on questions, activities, and assessments; provide thorough responses to questions and concerns raised throughout the course; respect diverse arguments from a variety of perspectives on the subject matter; make them aware of concerns I have with their performance or ability to succeed in the course; be available to them as a source of support in their learning.  If at any time I am forced to step away from the course for more than 24 hours, such as in the case of illness or personal emergency, I will notify the class as soon as possible and provide an additional point of contact for further information.

**Response Time and Feedback**

Whenever you submit a question by email or phone, you can expect feedback within 24 hours on a weekday and with 72 hours on a weekend.  Queries received on the last day of a weekend or a holiday break will be answered by the end of the day immediately following.  You can expect feedback on most assignments within 48 hours.  Where longer grading time is required, a timeframe for feedback will be shared with the assignment instructions or in a class-wide announcement.

**Disability Services**

To obtain disability-related academic accommodations students with documented disabilities must contact the course instructor and the Office of Student Access and Disability Services (OSA) as soon as possible. Truman complies with ADA requirements. For additional information, refer to the [Office of Student Access and Disability Services website](http://disabilityservices.truman.edu/) or contact by phone at (660) 785-4478 or [email](mailto:studentaccess@truman.edu).

# Title IX

Truman State University, in compliance with applicable laws and recognizing its deeper commitment to equity, diversity and inclusion which enhances accessibility and promotes excellence in all aspects of the Truman Experience, does not discriminate on the basis of age, color, disability, national origin, race, religion, retaliation, sex (including pregnancy), sexual orientation, or protected veteran status in its programs and activities, including employment, admissions, and educational programs and activities.   Faculty and staff are considered “mandated reporters” and therefore are required to report potential violations of the University’s Anti-Discrimination Policies to the Institutional Compliance Officer.

Title IX prohibits sex harassment, sexual assault, intimate partner violence, stalking and retaliation.  Truman State University encourages individuals who believe they may have been impacted by sexual or gender-based discrimination to consult with the Title IX Coordinator who is available to speak in depth about the resources and options.  Faculty and staff are considered “mandated reporters” and therefore are required to report potential incidents of sexual misconduct that they become aware of to the Title IX Coordinator.

For more information on discrimination or Title IX, or to file a complaint contact:

Institutional Compliance Officer, Title IX and Section 504 Coordinator  
Office of Institutional Compliance  
Violette Hall, Room 1308  
100 E. Normal Ave  
Kirksville, MO  63501  
Phone: (660) 785-4354  
[titleix@truman.edu](mailto:titleix@truman.edu)

Go online for the institution’s [complaint procedure](https://wp-internal.truman.edu/provost/files/2020/11/Truman-State-University-Complaint-Reporting-and-Resolution-Procedure9.2020.pdf) and the [complaint form](http://titleix.truman.edu/make-a-report/).

**Substantive Interaction**

Truman policy and federal regulations require that students demonstrate that they are academically engaged in the courses they take.  You must meet this requirement within the first week of the term, beginning at 12:00 am on Wednesday, August 18 and ending 11:59 pm Saturday August 24.  Failure to do so, or to provide an explanation of an extenuating circumstance by that date and time will result in your removal from the course.  Under certain circumstances, removal could impact your scholarship eligibility or financial aid.  For the purposes of this class, establishing academic engagement requires, at a minimum, introduce yourself in the discussion board.

**Academic Dishonesty**

The General Catalog states:

Students are expected to do their own academic work. Any student involved in cheating on a paper, an examination or in any other form of academic dishonesty is subject to disciplinary action, including suspension or expulsion from the class, the student’s academic program, or the University.

More information can be found in the [General Catalog](http://catalog.truman.edu/) and the [Student Conduct Code](http://policies.truman.edu/policylibrary/student-conduct-code/) Section 8.050.1.

In this class, work you do on individual assignments should be your own work, and work on group projects should reflect the work of group members. In general, you should have a sense for when you are acting unethically, and not do that. If you have questions, you should let me know.

Having said that, here is a non-exhaustive list of standards concerning common situations and questions that have come up in the past:

* **Study Groups:** I encourage you to exchange ideas through our Blackboard discussion forums or though study groups you might form with other class members. However, after you discuss an assignment, you should disengage from the group and make sure to write up your work on your own. This is the only way to make sure that **you** understand what’s going on!
* **Credit Where Credit is Due:** If you do work with a study group, give credit to the members of the group in your assignment write-up.
* **Outside Sources:** Especially when writing code, you don’t live in a vacuum. It’s natural to look online for solutions to coding problems. When you do this, or access other outside sources (living or non-living), you should also give credit. A link to the site where you found the code snippet you adapted, or an acknowledgment of the human who helped you, etc.

**Credit Hour Justification**

The minimum investment of time by the average Truman student necessary to achieve the learning goals in this course are not less than one hour (50 minutes) of classroom instruction and a minimum of two hours of out of class student work each week per credit hour awarded or at least the equivalent of three hours (2:50) of laboratory work, internships, practica, and other academic work each week per credit hour awarded. This average time per week for an average student may have weekly variations.

**FERPA**

Education records are protected by the [Family Education Right to Privacy Act (FERPA)](http://www.truman.edu/registrar/ferpa/).  As a result, course grades, assignments, advising records, etc. cannot be released to third parties without your permission.  There are, however, several exceptions about which you should be aware.  For example, education records can be disclosed to employees or offices at Truman who have an “educational need to know”.  These employees and offices may include your academic advisor, the Institutional Compliance Officer, the Registrar’s Office, or Student Affairs depending on the type of information