**Welcome:** to the final step in the Master's Program - the THESIS. It's our last chance to make sure you're ready to go out as a Professional Data Scientist and Analytic Storyteller!

This is a relatively unstructured experience where you create something meaningful that brings together your skills from multiple courses (and elsewhere) to really show off your skills, both with Data Science, and with Analytical Storytelling.

**Instructor:** Scott Alberts **Academic Success Mentor:** Ashley Harding

**Email:** [salberts@truman.edu](mailto:salberts@truman.edu)  **Office:** Pickler Library 108B

**Office Hours:** MW 9:30-11:30 am, **Phone**:660-785-7404

and Thursday 5:00-6:00 pm **Email:** [aharding@truman.edu](mailto:aharding@truman.edu)

**Class Meeting: Thursday 6:00-7:00 pm Availability:** Byappointment**,** M-Th

Zoom LInk for Office Hours and Class <https://truman-edu.zoom.us/j/708734757>

# **Communication:**

Plan on attending the weekly Zoom meeting, or let me know if that time doesn't work for you. Otherwise, the best way to contact me is through e-mail. All messages will be answered within 48 hours on weekdays and within 72 hours on weekends. I get a lot of e-mails, and many of them (not from you) are junk. I check email multiple times each day. I want to read yours first, so make sure your subject line starts with PDAT 630 and includes something descriptive, such as:

PDAT 630: Handout Formatting Questions

# Course Description: the culmination of the Masters in Data Science and Analytic Storytelling.

Students, working alone or in a pre-approved group, complete a data science project within their given discipline. It will involve a substantial written paper as well as an online presentation.

# **Prerequisites:** Successful completion of at least 24 credits in the PDAT program:

# Graduate School Requirements: Your completed thesis will be filed with the library and is available to future researchers online UMI ProQuest. Make sure you follow the rules for Electronic Thesis Submission as explained here:

<https://wp-internal.truman.edu/graduate/thesis-examining-committee-reports/>

Their rules are quite inflexible, so we MUST make their deadlines.

But, then you'll join this list: <http://library.truman.edu/archives/masterthesesindex.asp>

# Pre-Semester work: By the first day of the semester, we need to know/submit:

* The Title/Committee Approval form
* a one-page description of the project/research to be performed, including the form of the audience-focused artifacts that will be created as part of the project and final submission;
* a mostly complete list of data courses, RPackages, and other sources needed for the project

# Course Objectives: the successful student will:

* Complete the various paperwork needed to finish the Masters;
* Perform independent study in data science utilizing the tools and techniques learned in earlier PDAT courses and elsewhere. This usually involves use software (typically R with RMarkdown, but could also be python, spark, or most anything else) to:
  + acquire, clean, and join data sets;
  + perform appropriate exploration, visualization, modeling and analysis;
  + support, refute, or explore a thesis of interest.
* Produce written artifacts:
  + a quality technical report/white paper, written in IMRaD/similar format, that details the project, the process, and conclusions, at the level of fellow PDAT students;
  + an related artifact, more than an abstract, appropriate for use with a subject expert with limited technical knowledge;
  + a collection of final creative artifacts, appropriate for their intended audience(s), that achieve the thesis project goals. These artifacts may take many forms, depending on the project. For example, they may consist of a traditional written paper and summary materials appropriate to various audiences. Or, they may be an interactive online dashboard, with a technical report detailing its construction. Many types of creative artifacts are possible, but must be approved as part of the proposal process.
* Present a meaningful thesis defense of the project, including:
  + a short presentation (or opening) aimed at a non-technical audience
  + a high-quality presentation with sophisticated visualization, modeling, and analysis;
  + appropriate supporting visual aids (slides, handouts, and other takeaways).
* Respond to questions and concerns raised by the review committee and the public to create a final submission that includes written artifacts, visual aids, references and appendices.

# Technical Requirements, textbooks, and other materials:

Members of the class must have regular access to a reliable computer, a webcam/microphone, and the Internet. Try to create a workspace where you can work for a couple of hours in peace every day. It is a good idea to also locate a backup site (for instance, a library, coffee shop, or a friend's house) in case of technical failure.

The class uses the software used in earlier classes, including statistical software (typically R, but can also be Python, Hadoop/Spark, or others). As your project requires it, we can give you access to the "fire" server (which now runs RStudio Server and RShiny), and/or Amazon Web Service accounts.

There is no required textbook for this course. Students will find the materials from prior courses useful, and will need to consult discipline-specific supplementary resources.

# Course Outline

Each week, a submission/communication is due, along with an update.

| **Week** | **Progress:** All weekly assignments are due 11:59 pm on Wednesdays |
| --- | --- |
| **0 (1/11)** | Submit suggestions for members of your Thesis committee and your Title. Submit a Proposal Document suitable for sharing. |
| **1 (1/18)** | Complete online progress report, including R/Python packages to be used and a mostly complete list of data sources and other bibliographic sources. |
| **2 (1/25)** | Submit an online progress report, outlining the security and ethical components you expect to include, either in the main report or in an appendix. Almost no answer is wrong other than "there aren't any security or ethical components to my work." |
| **3 (2/1)** | Complete online progress report, detailing information about the technical, subject expert and non-technical audiences that will be the focus of the various products. |
| **4 (2/8)** | Complete online progress report. |
| **5 (2/15)** | Submit draft technical paper (to the instructor). Appendices don't have to be finished, but they should all be listed in the table of contents by now. |
| **6 (2/22)** | Complete online progress report, including more pieces of your packet  Somewhere around here, we'll be moving from Blackboard to a shared cloud folder, to allow committee members to see your materials. |
| **7 (3/1)** | For projects finishing during the current term, a "Public Draft" packet that is hopefully ready to send to the Thesis committee for review and feedback. |
| **8 (3/8)** | Complete online progress report, including any last pieces of the Final Draft. (for students finishing later, submit a longer online progress report including a proposed timeline for completion that varies from what is here) |
| ***3/15*** | *Truman Spring Break: Nothing due this week* |
| **9 (3/22)** | Incorporate committee feedback into the "Final Draft" to create the truly final paperwork, to allow scheduling of the Thesis presentation/defense. |
| **10 (3/29)** | Submit Revisions (multiple rounds if needed) |
| **11 (4/5)** | With instructor/committee approval, advertise presentations to STAT/PDAT. Practice presentation with peers and/or instructor/committee members |
| **12 (4/12)** | Submit Final Slides/handouts by 7 PM the day before the presentation. Deliver Oral Defense (including non-technical "public" presentation), either online or in person, no later than 4/14 (sooner is better) |
| **13 (4/19)** | Work on final edits and graduate school paperwork |
| **14 (4/26)** | Summer Finishers submit a final update.  All final artifacts are due to the graduate office by April 28th |
| **15 (5/5)** | Make arrangements to attend May 13 Commencement. |
| **Fin** | Data Science Hooding Ceremony is at 10 AM on 5/13/2022  Campus-wide Commencement is at 2 PM |

By the end of week 6, we should know whether this project is ready for presentation by week 14, or needs to continue into a future term. An "In-progress" grade is automatic for any projects making at least some progress, with a "Final Draft" to be submitted no later than week 9 for presentation in week 12, and a successful spring graduation..

If you do not complete the thesis within a year, you will need to enroll in an additional one-credit course, which may cause problems for financial aid, employers paying for your coursework, etc.

# Evaluation**:**

Successful completion of this course will usually result in successful completion of the Masters' Degree in Data Science and Analytic Storytelling. Two additional readers (beyond the instructor) do not provide formal grades, but must sign the final work for a passing grade to be assigned.

Grades will be based on the following assignments:

* Check-ins (at least 7, one each week that doesn't have an Assignment submitted),
* Assignments
  + proposal, Week 1,
  + draft technical paper submission, Week 4,
  + "public draft" of all artifacts, Week 6 (unless in-progress until fall)
* Final Artifacts:
  + final technical paper,
  + presentation(s), and
  + collected artifacts

Important submissions will be assigned an overall rubric score on a 1–5 scale, with others only graded for completion. The table below shows minimal requirements for each final grade. Notice that the difference between grades is not just a quality product at the end but ongoing communication and checkpoint submission.

| Grade | Check-ins and Assignments (7) | Artifact Rubric Score  (Technical Paper, Presentation(s), Artifact Collection) | Communication |
| --- | --- | --- | --- |
| A | All Submitted | 14/15 or 15/15 | Good Communication Throughout Class |
| B | At least 5 submitted | 11/15 or higher | Occasional Lapses in Communication |
| C | At least 3 submitted | 9/15 or higher | Minimal Communication |
| D | Not given |  | This course does not use a D grade |

An "In-Progress" grade is almost always a better choice than half-finishing a lousy Thesis project, even if it could earn a C. Still, if we delay your completion until Fall, we will want a plan in place to finish everything by the end of the fall semester.

# Work Expectations and Credit Hours:

As a three-credit graduate-level experience, this course expects a typical student will be successful if they spend approximately 120 total hours on the course (hours in higher education are 50 minutes long). This class (and any eight-week course) goes twice as fast as a normal 16-week semester class; you should plan to spend approximately 15-22 hours on this class EVERY week. Your actual time will vary as you progress through the project, including time spent before the term started and (if you take an "in-progress" grade), after the term is completed. It's recommended to spend an hour or two on the course most every day, rather than infrequent marathon sessions.

**Academic Integrity:** In this class, I treat you and your colleagues as burgeoning professional data scientists and statisticians. Statisticians and other scientists are expected to live up to a high ethical standard. Besides normal class expectations of avoiding cheating, lying, or plagiarizing, statisticians must ensure that they are fair to their clients, their subjects, and even their data itself. Poor citations are counted against your grade, but failing to acknowledge the work that was done by someone else is considered cheating. Cheating and lying is generally a poor choice and will result in an F in this course and referral for program-wide and campus-wide action. <http://www.amstat.org/ASA/Your-Career/Ethical-Guidelines-for-Statistical-Practice.aspx>

It is Truman's expectation that assignments submitted will be the original work of the student, using proper citation when building upon the work of others. When third-party code is included, its author and source must be properly cited, usually as a comment within the code itself.

Enrollment in this course indicates a student's desire or employment requirement to actually learn the material presented, with the assumption that the student will be expected to demonstrate a working knowledge of the associated skills in the workplace. Students who take the time to do the coursework themselves ensure that they are fully prepared to perform similar tasks in the workplace where they will be judged by their peers and supervisors. This is the place for students to make mistakes and learn to do it right themselves.

Students found to be in violation of this policy may be removed from the course, and may receive further sanctions, up to and including removal from the program and referral to the Office of Student Affairs for additional campus-wide sanctions.

**Health and Illness:** If you are sick, isolated, quarantined away from your computer, or otherwise unable to do your work, we can figure it out, but let the Instructor and Success Mentor know ASAP so that we can make arrangements before you fall too far behind.

**Campus Policies:** In addition to policies listed here in the syllabus, this class follows all campus policies, including those related to Disability Services, Emergency Procedures, Discrimination and Title IX listed here: <https://wp-internal.truman.edu/provost/syllabus-resources-for-students/>  
*(Note: I am a mandatory Reporter for Title IX and Discrimination Issues, but I can refer you to confidential counseling)*

**Substantive Interaction:** Truman policy and federal regulations require that students demonstrate that they are academically engaged in the courses they take. This policy is designed to protect students from being charged tuition for classes they never started nor intended to complete. Students must complete at least one Blackboard assignment or discussion post within the first week of the term. Failure to do this within the first week of class, or to provide an explanation of an extenuating circumstance, will result in your removal from the course. Under certain circumstances, removal from the course could impact your scholarship eligibility or financial aid.

**Feedback and Complaints:** If you have a problem or concern, please talk to the instructor or student success mentor early, so that we can solve the problem quickly and informally. If you are unhappy with us, informal feedback or formal complaints may be sent to:

~~Hyun-Joo Kim, Ph.D Data Science Program Director,~~ [~~hjkim@truman.edu~~](mailto:hjkim@truman.edu)~~;~~ (that's my wife)

~~K Scott Alberts, PhD, Chair, Department of Statistics,~~ [~~salberts@truman.edu~~](mailto:salberts@truman.edu)~~,~~ (that's me)

Tim Walston, Ph.D, Dean of Science & Mathematics at [samdean@truman.edu](mailto:samdean@truman.edu), 660-785-4248

Jeanne Harding, Director of Graduate Studies, at [gradstudies@truman.edu](mailto:gradstudies@truman.edu), 660-785-4109

Charles McAdams, EdD, Interim Provost, at [cmcadams@truman.edu](mailto:jquinzer@truman.edu), 660-785-4105

Sue Thomas, PhD, President, at [suethomas@truman.edu](mailto:suethomas@truman.edu), 660-785-4100