# DIG 6836: Humanistic Data Analysis

* **Instructor:** Dr. Rudy McDaniel
* **Email:** rudy@ucf.edu
* **Office:** Via Office (VAB-117a) Zoom; Monday 2:00 - 5:00 PM
* **Course Meeting:** Monday 6:00 - 7:50 PM

## Contents

## Course Description

Texts & Technology work is situated on an understanding of “new” media mechanisms, which includes the underlying layer of code. Throughout the core Texts & Technology courses, we work to bridge the gaps of text, technology, and theory. In *Humanistic Data Analysis*, we will put this understanding into practice, building towards confidence and flexibility in using procedural tools and digital methods in humanities research. While there are many languages potential relevant to T&T and the digital humanities broadly, this course focuses on Python, which is currently one of the top choices for data analysis thanks to its tools (including Jupyter Notebooks, which we will employ) and flexibility.

The main practical objective of this course is to provide graduate students with the necessary skills to utilize Python and associated tools for conducting research in the humanities. By the end of the course, students will be equipped with the knowledge and practical experience to analyze and interpret data within a humanistic context.

With this goal in mind, the bulk of the course will be spent experimenting. While at first it may seem there is minimal “writing” in this course, Annette Vee’s framework reminds us that code itself fits in the rubric of literacy, and through reading and writing code (and using it as a lens for seeing other “texts”) we will build that procedural literacy. This may mean stepping out of our comfort zones in order to remix content, experiment with strategies, fail often, perservere, and ultimately create projects that are useful to our goals as digitally literate scholars operating in an ever-evolving technological landscape.

At the end of the course, we will put this knowledge into use through a larger project, building upon the exercises towards a potential future research application in an area of interest.

Each week, plan on following instructions in the Webcourses module regarding readings and online exercises to accompany the in-class discussion. Each module will be divided into three sections:

## Student Learning Objectives

* Develop proficiency in Python programming and associated libraries/tools for data analysis in the humanities. This will include an understanding of programming, including basic terminology and the use of variables, functions, loops, libraries, and conditional statements.
* Understand the installation, configuration, syntax, and debugging methods required for Python.
* Apply data manipulation techniques to clean, preprocess, and transform diverse datasets commonly encountered in humanities research.
* Conduct exploratory data analysis to gain insights into patterns, trends, and distributions within humanities datasets.
* Employ text processing and natural language processing techniques to analyze and extract meaningful information from textual data in the humanities.
* Utilize web scraping and data collection methods to gather relevant data from online sources for humanities research.
* Learn about and apply appropriate programmatic techniques and/or algorithms unique to your individual research question(s) (e.g., geospatial analysis, network analysis, machine learning, topic modeling, time series analysis, text mining, etc.).
* Create visually appealing and informative data visualizations for communicating research findings in the humanities.
* Demonstrate an understanding of ethical considerations and responsible practices in humanistic data analysis, including privacy, security, and data protection.
* Reflect upon the design of technology, with critical attention to its engagement with theoretical frameworks and the assumptions and dangers of algorithmic methods.
* Design and execute an independent research project that applies Python and associated tools to analyze and interpret data within a humanistic context.
* Communicate research findings effectively through written reports and oral presentations.

These learning outcomes aim to ensure that students acquire both the technical skills necessary for data analysis in the humanities and a strong understanding of the ethical considerations and research practices relevant to the field.

## Course Structure

* **Weekly Readings and Lecture.** The full schedule of required readings is listed in the syllabus: additional recommended readings will be provided in each module. Weekly lectures for this course will be pre-recorded and available online.
* **In-Person Activities.** This course follows the flipped classroom approach: lectures are online, and each week’s in-person meeting will involve a combination of hands-on design and coding exercises, lab collaboration time, and discussions focused on understanding and applying concepts from the readings. All activities will be documented and available online for those unable to attend
* **Coding Exercises.** A weekly online assignment to build your skills in programming, debugging, and applying code solutions to research and design problems.

## Materials and Texts

This course requires a mix of applied and theoretical readings, including some open access materials. The primary texts include:

* Drucker, Johanna. 2023. [*The Digital Humanities Coursebook: An Introduction to Digital Methods for Research and Scholarship.*](https://go.openathens.net/redirector/ucf.edu?url=https%3A%2F%2Fwww.taylorfrancis.com%2Fbooks%2Fmono%2F10.4324%2F9781003106531) Routledge. (This text is available as an unlimited access ebook through UCF library: follow the link and log in to download it.)
* Karsdorp, Folgert, Mike Kestemont, and Allen Riddell. 2023. *Humanities Data Analysis: Case Studies with Python.* Princeton University Press.

Supplementary suggested readings are included in each module to guide additional exploration of programming methods and concepts explored.

## Evaluation and Grading

| Points | Assignment Summary | Due Date |
| --- | --- | --- |
| 5 | **Activity Verification** - Complete the brief survey posted on Webcourses as soon as possible to confirm your enrollment in the course. As this is required by the university, please attend to it as soon as possible at the start of classes. | Friday, August 25, 2023 |
| 75 | **Coding Exercises** - Weekly discussions will consist of sharing preliminary readings and arguments, as well as progress towards the final paper. Students are encouraged to use screenshots from the works under discussion, and models will be provided for annotating and engaging through these visuals. | Weekly, 11:59pm each Sunday |
| 20 | **Reflection** - During the final exam week, students will complete a written reflective essay on their journey, with particular consideration to next steps and potential future applications of programming tools introduced throughout the semester. | Monday, December 4, 2023 |

Students can access their grades and feedback at any time using the Grade Book function of Webcourses. All assignments will be submitted through Webcourses. Plan on checking the site at least twice a week for updates and assignment information. Grades are calculated out of 100 following a standard letter scale.

Late work is accepted without penalty through the next major assignment deadline. If circumstances require extension beyond that deadline, please reach out to the instructor immediately. As assignments throughout the course are designed to build on the previous exercise, assignments must be completed in sequence.

There is no extra credit work available in this class. Grades will be available through Webcourses and updated weekly.

## Mixed Mode Course Structure

This course uses a mixed mode format, and relies upon students to complete all readings, engage with online videos, and join in on course discussions. All assignments are due at the close of their listed module, but will be accepted with no penalty through the next listed deadline. Once an assignment closes, late work will not be accepted unless an additional extension has already been approved by the instructor: please reach out early if circumstances will require additional time!

* The course will meet at the scheduled time on campus unless otherwise noted in the weekly module.
* Office hour assistance is additionally available both through text on Webcourses messages and via Zoom: Zoom is recommended for advanced technical problems, where screen-sharing might be helpful to resolving errors.
* Students will need access to a reliable internet connection and computer to participate in this course. Due to the Python installation needs, administrative access to the system is required to complete assignments.
* In the event of an emergency or medical challenge, additional flexibility beyond the grading guidelines is available: when anticipated, students should reach out to the instructor as soon as feasible to form a plan or discuss an incomplete if needed.

## Weekly Schedule

### Week One: Gearing Up (Monday, August 21, 2023)

We start the semester by configuring our systems for Python: we’ll work through the installation process and get to know Python, Thonny, Google CoLab, and Git as the tools we’ll rely upon throughout the semester.

* *Digital Humanities Coursebook*: Chapter 1
* **Due: Activity Verification (Friday August 27th)**
* **Coding Exercise One: Python Pals**

### Week Two: Data (Monday, August 28, 2023)

As we test out our configurations, we will explore the different types of data that T&T work can engage, as well as start to consider the practicalities of marking, cleaning, and using data.

* Turkel, William J. and Adam Krymble. [Code Reuse and Modularity in Python.](https://programminghistorian.org/en/lessons/code-reuse-and-modularity)
* Turkel, William J. and Adam Krymble. [Downloading Web Pages in Python.](https://programminghistorian.org/en/lessons/working-with-web-pages)
* *Digital Humanities Coursebook*: Chapter 2
* **Coding Exercise Two: Web Content**

### Week Three: Strings (Holiday - Online Module)

We’ll build our comfort with the language of programming through exploring building blocks, with an emphasis on strings as the foundation of textual data, as well as variables, functions, and loops.

* Turkel, William J. and Adam Krymble. [Manipulating Strings in Python.](https://programminghistorian.org/en/lessons/manipulating-strings-in-python)
* Turkel, William J. and Adam Krymble. [From HTML to List of Strings.](https://programminghistorian.org/en/lessons/from-html-to-list-of-words-1)
* *Digital Humanities Coursebook*: Chapter 3
* **Coding Exercise Three: Strings**

### Week Four: Metadata (Monday, September 11, 2023)

Through thinking about both the texts we engage and their metadata, markup, and descriptions, we’ll engage both the pragmatic challenges of standards (XML, TEI, JSON, and so forth) and the research possibilities.

* Turkel, William J. and Adam Krymble. [Normalizing Textual Data with Python.](https://programminghistorian.org/en/lessons/normalizing-data)
* Turkel, William J. and Adam Krymble. [Counting Word Frequency with Python.](https://programminghistorian.org/en/lessons/counting-frequencies)
* *Digital Humanities Coursebook*: Chapter 4
* **Coding Exercise Four: Counting**

### Week Five: Jupyter (Monday, September 18, 2023)

As we level up our methodologies, we’ll start to explore the capacity of Jupyter Notebooks, with particular attention to how we can start to build reusable, structured methodologies for working in Python.

* Dombrowsi, Quinn, Tassie Gniady, and David Kloster. [Introduction to Jupyter Notebooks](https://programminghistorian.org/en/lessons/jupyter-notebooks)
* *Humanities Data Analysis*: Chapter 1
* **Coding Exercise Five: Jupyter Configuration**

### Week Six: Data Structures (Monday, September 25, 2023)

Working with larger datasets requires more intentional structures: we’ll explore databases and build our comfort with manipulating and parsing CSV, PDFs, JSON, XML, and HTML documents.

* *Humanities Data Analysis*: Chapter 2
* *Digital Humanities Coursebook*: Chapter 5
* **Coding Exercise Six: Structured Data**

### Week Seven: Vectors (Monday, October 2, 2023)

Through leveling up our text processing, we’ll build our confidence in another data structure, arrays, while starting to think about visual as well as textual dimensions to exploring data.

* *Humanities Data Analysis*: Chapter 3
* *Digital Humanities Coursebook*: Chapter 6
* **Coding Exercise Seven: Initial Analysis**

### Week Eight: Analysis (Monday, October 9, 2023)

By starting to combine humanist theoretical lenses with our new technical methodologies, we’ll engage in thoughtful critique of cultural data “mining” and its challenges.

* *Humanities Data Analysis*: Chapter 4
* *Digital Humanities Coursebook*: Chapter 7
* **Coding Exercise Eight: Cultural Analysis?**

### Week Nine: Numbers (Monday, October 16, 2023)

While the quantitative is rarely centered in Texts & Technology, some measurable datasets can have revealing numeric aspects. We’ll explore methods for visualizing statistics and probability in humanities data.

* *Humanities Data Analysis*: Chapter 5 and 6
* **Coding Exercise Nine: Numbers**

### Week Ten: Maps (Monday, October 23, 2023)

Drawing on models of GIS and spatial humanities, we’ll explore the potential for data maps and plots across humanities datasets, thinking through both traditional and thematic approaches to mapping.

* *Humanities Data Analysis*: Chapter 7
* *Digital Humanities Coursebook*: Chapter 8
* **Coding Exercise Ten: Mapping**

### Week Eleven: Voices (Monday, October 30, 2023)

While understandably subject to critique, computational methods for attributing authorship and finding patterns in writing styles have attracted a lot of attention. We’ll explore these mechanisms, with attention to the theorizing of “voice.”

* *Humanities Data Analysis*: Chapter 8
* *Digital Humanities Coursebook*: Chapter 9
* **Coding Exercise Eleven: Principal Component Analysis**

### Week Twelve: Models (Monday, November 6, 2023)

As we continue to critique and explore distant reading, we will engage with methods of topic modeling, and consider the potential and challenges of integrating these approaches alongside humanities approaches.

* *Humanities Data Analysis*: Chapter 9
* *Digital Humanities Coursebook*: Chapter 10
* **Coding Exercise Twelve: Texts, Three Ways**

### Week Thirteen: Good Enough (Monday, November 13, 2023)

Drawing on the models and questions from our texts, we’ll consider the approach of “good enough,” and discuss how we can bring these lessons to our projects, particularly in terms of allowing for failure and iteration. We’ll also explore methods for translating knowledge from Python to other programming languages.

* *Humanities Data Analysis*: Epilogue
* *Digital Humanities Coursebook*: Chapter 11
* **Coding Exercise Thirteen: Interface**

### Week Fourteen: Critiques (Monday, November 20, 2023)

Using both *Race After Technology* and the work of other scholars from our core list, we will critique the methods we’ve put into place, considering how we might reshape and reimagine our technologies and code in light of these challenges.

* *Digital Humanities Coursebook*: Chapter 12
* **Coding Exercise Fourteen: Design**

### Week Fifteen: Design Futures (Monday, November 27, 2023)

During the final project showcase, we will embrace both the successes and failures of the process of integrating Python-driven methods and tools into our scholarship, and brainstorm next steps.

* *Digital Humanities Coursebook*: Coda
* **Coding Exercise Fourteen Due**

### Final Exam/Reflection (Monday, December 4, 2023)

## General Policies

### COVID-19 Statement for Fall 2023

**Notifications in Case of Changes to Course Modality** If the instructor falls ill during the semester, there may be temporary changes to this course, including having a backup instructor take over the course or going remote for a short time. Please look for announcements or mail in Webcourses@UCF or Knights email for any temporary alterations to this course.

Students should contact their instructor(s) as soon as possible if they miss class for any illness to discuss reasonable adjustments that might need to be made. When possible, students should contact their instructor(s) before missing class.

### Academic Integrity

[The Center for Academic Integrity (CAI)](http://academicintegrity.org/) defines academic integrity as a commitment, even in the face of adversity, to five fundamental values: honesty, trust, fairness, respect, and responsibility. From these values flow principles of behavior that enable academic communities to translate ideals into action.

**UCF Creed:** Integrity, scholarship, community, creativity, and excellence are the core values that guide our conduct, performance, and decisions.

1. Integrity: I will practice and defend academic and personal honesty.
2. Scholarship: I will cherish and honor learning as a fundamental purpose of my membership in the UCF community.
3. Community: I will promote an open and supportive campus environment by respecting the rights and contributions of every individual.
4. Creativity: I will use my talents to enrich the human experience.
5. Excellence: I will strive toward the highest standards of performance in any endeavor I undertake.

Students should familiarize themselves with [UCF’s Rules of Conduct](https://scai.sdes.ucf.edu/student-rules-of-conduct/). According to Section 1, “Academic Misconduct,” students are prohibited from engaging in

1. Unauthorized assistance: Using or attempting to use unauthorized materials, information or study aids in any academic exercise unless specifically authorized by the instructor of record. The unauthorized possession of examination or course-related material also constitutes cheating.
2. Communication to another through written, visual, electronic, or oral means: The presentation of material which has not been studied or learned, but rather was obtained through someone else’s efforts and used as part of an examination, course assignment, or project.
3. Commercial Use of Academic Material: Selling of course material to another person, student, and/or uploading course material to a third-party vendor without authorization or without the express written permission of the university and the instructor. Course materials include but are not limited to class notes, Instructor’s PowerPoints, course syllabi, tests, quizzes, labs, instruction sheets, homework, study guides, handouts, etc.
4. Falsifying or misrepresenting the student’s own academic work.
5. Plagiarism: Using or appropriating another’s work without any indication of the source, thereby attempting to convey the impression that such work is the student’s own.
6. Multiple Submissions: Submitting the same academic work for credit more than once without the express written permission of the instructor.
7. Helping another violate academic behavior standards.
8. Soliciting assistance with academic coursework and/or degree requirements.

### Responses to Academic Dishonesty, Plagiarism, or Cheating

Students should also familiarize themselves with the procedures for academic misconduct in UCF’s student handbook, [The Golden Rule](https://goldenrule.sdes.ucf.edu/). UCF faculty members have a responsibility for students’ education and the value of a UCF degree, and so seek to prevent unethical behavior and respond to academic misconduct when necessary. Penalties for violating rules, policies, and instructions within this course can range from a zero on the exercise to an “F” letter grade in the course. In addition, an Academic Misconduct report could be filed with the Office of Student Conduct, which could lead to disciplinary warning, disciplinary probation, or deferred suspension or separation from the University through suspension, dismissal, or expulsion with the addition of a “Z” designation on one’s transcript.

### Course Accessibility Statement

This course is built with universal design for learning principles in mind: if you encounter challenges with any of the course materials, assignments, platforms, or requirements, please reach out for assistance, and know that additional support is always available regardles of documentation. If changes in course modality occur and you require additional accomodations, please reach out as soon as possible.

Additionally, the University of Central Florida is committed to providing access and inclusion for all persons with disabilities. Students with disabilities who need access to course content due to course design limitations should contact the professor as soon as possible. Students should also connect with [Student Accessibility Services (SAS)](http://sas.sdes.ucf.edu/) (Ferrell Commons 185, sas@ucf.edu, phone 407-823-2371). For students connected with SAS, a Course Accessibility Letter may be created and sent to professors, which informs faculty of potential course access and accommodations that might be necessary and reasonable. Determining reasonable access and accommodations requires consideration of the course design, course learning objectives and the individual academic and course barriers experienced by the student. Further conversation with SAS, faculty and the student may be warranted to ensure an accessible course experience.

### Campus Safety Statement

Emergencies on campus are rare, but if one should arise during class, everyone needs to work together. Students should be aware of their surroundings and familiar with some basic safety and security concepts.

* In case of an emergency, dial 911 for assistance.
* Every UCF classroom contains an emergency procedure guide posted on a wall near the door. Students should make a note of the guide’s physical location and review the [online version](http://emergency.ucf.edu/emergency_guide.html).
* Students should know the evacuation routes from each of their classrooms and have a plan for finding safety in case of an emergency.
* If there is a medical emergency during class, students may need to access a first-aid kit or AED (Automated External Defibrillator). To learn where those are located, see [locations](https://ehs.ucf.edu/automated-external-defibrillator-aed-locations).
* To stay informed about emergency situations, students can sign up to receive UCF text alerts by going to [MyUCF](https://my.ucf.edu) and logging in. Click on “Student Self Service” located on the left side of the screen in the toolbar, scroll down to the blue “Personal Information” heading on the Student Center screen, click on “UCF Alert”, fill out the information, including e-mail address, cell phone number, and cell phone provider, click “Apply” to save the changes, and then click “OK.”
* Students with special needs related to emergency situations should speak with their instructors outside of class.
* To learn about how to manage an active-shooter situation on campus or elsewhere, consider viewing this [video](https://youtu.be/NIKYajEx4pk).

### Deployed Active Duty Military Students

Students who are deployed active duty military and/or National Guard personnel and require accommodation should contact their instructors as soon as possible after the semester begins and/or after they receive notification of deployment to make related arrangements.

### Authorized Events and Religious Observances

Students who represent the university in an authorized event or activity (for example, student-athletes) and who are unable to meet a course deadline due to a conflict with that event must provide the instructor with documentation in advance to arrange a make-up. No penalty will be applied.

Students must notify their instructor in advance if they intend to miss class for a religious observance. For more information, see the [UCF policy](http://regulations.ucf.edu/chapter5/documents/5.020ReligiousObservancesFINALJan19.pdf).