**Data: A User Manual**

HIST 4261/5261

History Department, University of Colorado Denver

Fall 2021 Professor Cameron Blevins

Monday/Wednesday 12:30-1:45 p.m. Office: 3107 Student Commons

Course location: TBD (Denver Campus) Email: cameron.blevins@ucdenver.edu

Office Hours: TBD Phone: 303-315-1776

**I. Overview**

**Catalog Description:**

Teaches the technical skills of data collection, processing, analysis, and visualization, along with the history and ethics of how societies, corporations, and governments have used and abused data over time.

**Course Description:**

Data is one of the most valuable commodities of the 21st century. In this course, students will learn the technical skills of data analysis along with the history of how data has been collected, used, and misused over time. First, students will learn how to ***work with data*** by creating, collecting, processing, cleaning, analyzing, and visualizing datasets using the R programming language. Second, students will learn about the longer ***history of data*** in order to understand the cultural and ethical dimensions of this topic. While learning how to write computer code students will also be examining historical questions such as: How has the US government gathered information about its citizens, from the first national census to 20th-century spy agencies? What is the relationship between 19th-century accounting practices and the rise of plantation slavery? Together, these two approaches will provide students with a “user manual” for how to effectively and ethically use data today.

**Learning Goals and Objectives:**

In this class, students will:

* Learn how to collect, process, analyze, and visualize data using the R programming language for statistical computing.
* Follow major historical shifts in how societies, corporations, and governments have used data over time.
* Understand the ways in which data is shaped by larger cultural contexts and power relations.
* Learn how data can cause harm while developing best practices for gathering, analyzing, and deploying data in ethical ways.
* Graduate Students: Understand the historiography of data studies and some of the major scholarly currents on this topic over the past half century
* Graduate Students: Gain teaching experience by leading class discussions and synthesizing feedback on student projects

*Note: This course provides 3.0 approved credit hours for the Undergraduate and Graduate Certificate in Digital Studies:* [*https://clas.ucdenver.edu/digital-studies-certificates/*](https://clas.ucdenver.edu/digital-studies-certificates/)*.*

**II. Course Materials**

The following two textbooks are available online as ebooks through Auraria Library. All other readings will be available online for free or posted to Slack.

* Taylor Arnold and Lauren Tilton, *Humanities Data in R: Exploring Networks, Geospatial Data, Images, and Text* (Springer, 2015)
* Catherine D'Ignazio and Lauren F. Klein, *Data Feminism* (MIT Press, 2020)

**III. Assignments and Evaluation**

**Participation (30%)**

You are expected to attend each class having completed the week’s readings and/or tutorials and be prepared to discuss them. Participation is based on a mix of several components:

1. **Attendance**: Attendance is crucial for you to learn and excel in this course and for maintaining a sense of community with your classmates. You are allowed three absences, no questions asked (roughly 10% of the class meetings). If, however, you run into unforeseen health, work, other challenges and need to miss additional classes, please email me and we will come up with a strategy for getting you caught up.
2. **In-Class Discussion**: Prior to class, you are expected to have completed the readings or other assignments and be prepared to discuss them with your classmates in class.
3. **Slack Posts**: As part of that preparation, each student will post to the course’s #discussion channel on Slack by 9AM the day of class. These can take the form of a series of questions, comments, GIFs, memes, etc. that reflect on the week’s readings. Please read each of your classmates’ posts prior to the start of class.
4. **Project Presentation**: Each class, one or two students will present a short overview of a data-driven project related to the week’s topic. You will provide a SHORT introduction to the project (3-5 minutes) and how it relates to the week’s theme or readings. You are welcome to discuss its background, strengths or shortcomings, etc. but keep in mind that you do not have very much time. Plan out ahead of time exactly what you need to convey to your classmates so that they can understand the project and its relationship to the theme. The goal of this assignment is two-fold. First, it will provide the class with a series of real-world examples of projects “in action.” Second, in trying to find your project you will get exposed to a wider body of data-driven research. Please email or Slack DM me with the project you are thinking of doing at least two days before class so that I can approve it. You will sign up for a presentation time during the first week of the semester.
5. **Lesson Plan *(Graduate Students only)***: In addition to your project presentation, you will make and then teach a lesson plan for the day’s major themes and readings. This can take any form you like (series of discussion questions, hands-on activity, etc.) but must be submitted to me at least 2 days before class.

**Coding Exercises (30%)**

You will be completing a series of seven coding exercises in order to help you learn how to work with data, primarily using the R programming language. Each exercise will focus on a particular dataset and ask you to apply the skills you learned in the preceding classes to analyze that dataset in some way. Unless otherwise noted, these will be submitted as a Jupyter Notebook (we will go over how to make and submit these at the beginning of the semester). Each exercise is worth 5% of your final grade and I will be dropping the worst score from your seven exercises.

* Graduate Students will need to complete two additional questions/problems included at the end of each of the coding exercises.

*See Course Schedule for Due Dates*

**History of Data Paper (15%)**

For this assignment you will write a 5-7 page paper (double-spaced) on a particular kind of data and its history (ex. the history of presidential election polling in the 20th-century United States). Some questions to consider: How has this data been collected, analyzed, or used? What kinds of cultural, social, and ethical dimensions surrounded its use? How has it changed over time? How is that kind of data used today and/or what are some lessons we might draw from this example in thinking about data in the present? This paper will require you to find and synthesize existing scholarship on this topic and should focus on an example of data that is *not* already covered in the class readings.

* Graduate Students: This paper should take the form of a 7-9 page (double-spaced) historiographical review essay that examines 3-4 major scholarly works on your topic.

*Due Sunday, Sunday November 7th by 11:59PM*

**Final Project (25%)**

In your final project you will find or create some kind of dataset, process it, analyze it, and then present an argument, interpretation, or narrative based on that data. This will take the form of Jupyter Notebook, and include the following components:

* Introduction: the questions you sought to answer about your data; any context that you think your reader will need to understand the data, including existing work on the topic; methodological discussion of how the data was collected and processed
* Main analysis of the dataset: a clearly communicated argument, interpretation, or narrative supported by compelling evidence and examples; appropriate visualizations of your dataset;
* Conclusion: the larger cultural context of this dataset; ethical concerns or issues around its collection or application
* Bibliography: list of existing scholarship and sources of your data
* Appendix: the code that you wrote, clearly organized and annotated for your reader to understand

Graduate students: As part of your final project grade you will be synthesizing feedback for *one* additional project, drawing on verbal feedback from the class during presentations and your own in-depth analysis of the project. This should take the form of a 1-2 page memo detailing the project’s strengths, weaknesses, and concrete suggestions for improvement.

* *Rough Draft due Sunday, December 5th by 11:59PM*
* *Presentations of final project in class December 6th and 8th*
* *Graduate Students: written feedback on your classmate’s project due Thursday, December 9th by 11:59PM*
* *Final Project due Sunday, December 12th by 11:59PM*

**V. Weekly Topics and Assignments**

Mon., Aug. 23: Introduction to the Course

Wed., Aug. 25: Getting Your Feet Wet

* “The Age of the Algorithm” 99% Invisible podcast (September 5, 2017).
* Install R and RStudio on your laptop
* Annotate the Course Syllabus on Google Docs

Mon., Aug. 30: Defining Data

* “Introduction: Why Data Science Needs Feminism” and “Chapter 1: The Power Chapter” in *Data Feminism* (48 pp.)
* Graduate: Julia Flanders and Fotis Jannidis, “Data modeling in a digital humanities context: an introduction” in Flanders and Jannidis, *The Shape of Data in Digital Humanities* (23 pp.)

Wed., Sep. 1: The Humble Spreadsheet

* Robert Kosara, “Spreadsheet Thinking Versus Database Thinking” https://eagereyes.org/basics/spreadsheet-thinking-vs-database-thinking
* Sandhya Kambhampati, “Cleaner, Smarter Spreadsheets Start with Structure” https://source.opennews.org/articles/building-cleaner-smarter-spreadsheets/

Mon., Sep. 6: No Class (Holiday)

Wed., Sep. 8: Data Structures in R

* “2.2 Calculator and Objects,” “2.3 Numeric Vectors,” “2.4 Logical Vectors” and “2.6 Character Vectors” in *Humanities Data in R*
* Brooke Foucault Welles, “On Minorities and Outliers: The Case for Making Big Data Small,” *Big Data & Society* 1, no. 1 (April 1, 2014) (2 pp.)

*Coding Exercise #1 Due Friday, 9/10 by 11:59PM*

Mon., Sep. 13: Slavery and Data I

* “Chapter 2: Collect, Analyze, Imagine, Teach” in *Data Feminism* (23 pp.)
* Jessica Marie Johnson, “Markup Bodies: Black [Life] Studies and Slavery [Death] Studies at the Digital Crossroads,” *Social Text* 36, no. 4 (137) (2018): 57–79. (22 pp.)
* Graduate: “Introduction” in Michel-Rolph Trouillot, *Silencing the Past* (1997) (30 pp.)

Wed., Sep. 15: Manipulating Vectors

* “2.5 Subsetting” in *Humanities Data in R*
* Graduate: Julia Flanders and Fotis Jannidis, “Data modeling in a digital humanities context: an introduction” in Flanders and Jannidis, *The Shape of Data in Digital Humanities* (23 pp.)

*Coding Exercise #2 Due Friday, 9/17 by 11:59PM*

Mon., Sep. 20: Slavery and Data II

* “Chapter 3: On Rational, Scientific, Objective Viewpoints from Mythical, Imaginary, Impossible Standpoints” in *Data Feminism* (23 pp.)
* “Chapter 3: Slavery’s Scientific Management” in Caitlin Rosenthal, *Accounting for Slavery:* *Masters and Management* (2018) (35 pp.)
* Graduate: “Chapter 4: Human Capital” in Caitlin Rosenthal, *Accounting for Slavery:* *Masters and Management* (2018) (36 pp.)

Wed., Sep. 22: Two-Dimensional Data

* “2.7 Matrices and Data Frames”, “2.8 Data I/O” in *Humanities Data in R*

*Coding Exercise #3 Due Friday, 9/24 by 11:59PM*

Mon., Sep. 27: Slavery and Data III

* “Introduction: The Value of Life and Death” and Chapter 6: Postmortem: Death and Ghost Values” in Daina Ramey Berry, *The Price for Their Pound of Flesh: The Value of the Enslaved, from Womb to Grave* (54 pp.)
* Graduate: “Chapter 4: Slavery and the Rise of Statistical Cartography” in Susan Schulten, *Mapping the Nation* (35 pp.)

Wed., Sep. 29: Manipulating Matrices and Dataframes

* “2.9 Advanced Subsetting” in *Humanities Data in R*

*Coding Exercise #4 Due Friday, 10/1 by 11:59PM*

Mon., Oct. 4: Government Data I

* “Chapter 4: What Gets Counted Counts” in *Data Feminism* (27 pp.)
* “Introduction” and “Chapter 1: Nature and Space” in James Scott, *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed* (49 pp.)
	+ Graduate: “Chapter 2: Cities, People, and Language” in Scott, *Seeing Like a State* (31 pp.)

Wed., Oct. 6: Plotting and Visualization

* “EDA I: Continuous and Categorical Data” and “EDA II: Multivariate Analysis” (Entire Chapters) in *Humanities Data in R*

Mon., Oct. 11: Government Data II

* “Chapter 4: The Census and Industrial America in the Gilded Age” and “Chapter 5: Building the Federal Statistical System in the Early Twentieth Century” in Margo Anderson, *The American Census: A Social History* (44 pp.)
* Graduate: “Chapter 5: The Cartographic Consolidation of America” in Susan Schulten, *Mapping the Nation* (39 pp.)

Wed., Oct. 13: Text Analysis

* Chapter 9. Natural Language Processing and Chapter 10. Text Analysis (Entire Chapters) in *Humanities Data in R*

*Coding Exercise #5 Due Friday, 10/15 by 11:59PM*

Mon., Oct. 18: Government Data III

* “Chapter 6: The Numbers Don’t Speak for Themselves” in *Data Feminism* (22 pp.)
* Heidi Tworek, “Communicable Disease: Information, Health, and Globalization in the Interwar Period” *American Historical Review* (2019) (29 pp.)
* Graduate: “Chapter 7: Counting the Unemployed and the Crisis of the Great Depression” in Margo Anderson, *The American Census: A Social History* (30 pp.)

Wed., Oct. 20: Mapping

* Chapter 7. Geospatial Data (Entire Chapter) in *Humanities Data in R*

Mon., Oct. 25: Corporate Data I

* Sharon Ann Murphy, “Selecting Risks in an Anonymous World: The Agency System for Life Insurance in Antebellum America,” *Business History Review* 82, no. 1 (April 2008): 1–30.
* “Chapter 1: Technologies of Publicity” in Sara Igo, *The Known Citizen: A History of Privacy in Modern America* (42 pp.)

Wed., Oct. 27: Making Data Machine-Readable

* Hadley Wickham, “Tidy Data” *Journal of Statistical Software*, Vol. 59, Issue 10 (2014). (23 pp.)
* Katie Rawson and Trevor Muñoz, “Against Cleaning” in *Debates in Digital Humanities 2019* (12 pp.)

*Coding Exercise #6 Due Friday, 10/29 by 11:59PM*

Mon., Nov. 1: Corporate Data II

* “Chapter 7: Show Your Work” in *Data Feminism* (29 pp.)
* “Chapter 8: Stories of One’s Self” in Sara Igo, *The Known Citizen: A History of Privacy in Modern America* (42 pp.)

Wed., Nov. 3: Adding Context to Data

* Timnit Gebru et al. "Datasheets for Datasets" <https://arxiv.org/abs/1803.09010> (24 pp.)

*History of Data Paper Due Sunday, 11/7 by 11:59PM*

Mon., Nov. 8: Corporate Data III

* Chapter 22-25 in Margaret O’Mara, *The Code: Silicon Valley and the Remaking of America* (48 pp.)

Wed., Nov. 10: Networks

* Chapter 6. Networks (Entire Chapter) in *Humanities Data in R*

Mon., Nov. 15: Data Today

* “Chapter 5: Unicorns, Janitors, Ninjas, Wizards, and Rock Stars” in *Data Feminism* (23 pp.)
* “Chapter 6: The Future of Information Culture” in Safiya Noble, *Algorithms of Oppression* (17 pp.)

Wed., Nov. 17: Image Analysis

* Chapter 8. Images (Entire Chapter) in *Humanities Data in R*

*Coding Exercise #7 Due Friday, 11/19 by 11:59PM*

Mon., Nov. 22: No Class (Holiday)

Wed., Nov. 24: No Class (Holiday)

Mon., Nov. 29: Student Choice

* Historical readings on topic chosen by students

Wed., Dec. 1: Final Project Workshop

* Technical readings on topic chosen by students

*Rough Draft of Final project due Sunday, December 5th by 11:59PM*

Mon., Dec. 6: Final Project Presentations

Wed., Dec. 8: Final Project Presentations

*Final Projects Due Sunday, December 12th by 11:59PM*

**VI. Course Policies**

**Attendance:** You are allowed three absences, no questions asked (roughly 10% of the class meetings), after which I will begin deducting points from your participation grade. If, however, you run into unforeseen health, work, other challenges and need to miss additional classes, please email me and we will come up with a strategy for getting you caught up.

**Late Assignments:** If unforeseen circumstances come up during the semester and you need an extension on a particular assignment, please email me. Otherwise you are expected to submit all assignments by the due date. Late papers will be penalized one-third of a grade for each day late.

**Syllabus Changes:** All dates and assignments in this syllabus are subject to modification. Any changes will be announced during the semester in class and over email/Canvas. If you miss a class meeting, you are responsible for contacting me or another student regarding any missed announcements.

**Classroom Behavior**: Treat your fellow classmates and professor with respect: listening attentively, not interrupting, and maintaining a civil discourse. Bullying or personal attacks will not be tolerated. Learn each other’s names and preferred pronouns, and always assume the best intentions in each other. If you have any issues, please talk to me directly so that I can address them.

**Grading Policies:**

All assignments will be graded according to the following grading scale:

94-100 A

90-93 A-

87-89 B+

84-86 B

80-83 B-

77-79 C+

74-76 C

70-73 C-

67-69 D+

64-66 D

60-63 D-

0-59 F

**Access, Disability, Communication:** I am committed to providing everyone the support and services needed to participate in this course. If you have a learning disability, I encourage you to notify me and to contact [Disability Resources and and Services](https://www.ucdenver.edu/student-services/resources/disability-resources-services/accommodations/Pages/accessing-accommodations.aspx) at 303-315-3510 or by e-mail at disabilityresources@ucdenver.edu. If you have a temporary medical condition or injury, please reach out to Disability Resources and Services as well.

**Health and Wellness:** Every student should prioritize their physical, emotional, and mental health. Any student who is struggling to do so (lacking a safe and stable place to live, having difficulty affording sufficient food to eat every day, facing financial stresses, etc.) is urged to consult the resources listed here: <https://www.ucdenver.edu/student/health-wellness>, and in particular the University’s [Counseling Center](https://www.ucdenver.edu/counseling-center).

**Students Called for Military Duty:** Students in the military with the potential of being called to military service and/or training during the course of the semester are encouraged to notify their school/college Associate Dean or Advising Office.

**Religious Observances**If, because of religious obligations, you have conflicts with required assignments/attendance please email me directly as soon as possible at the beginning of the semester.

**CLAS Incomplete Policy:** The faculty in the College of Liberal Arts and Sciences passed the following policy relating to the awarding of incomplete grades:

Incomplete grades (IW or IF) are NOT granted for low academic performance. To be eligible for an incomplete grade, students MUST:

* Have participated in the class for a significant proportion of the term
* Have successfully completed a significant proportion of the course assignments
* Have special circumstances (verification may be required) that preclude the student from attending class and/or completing graded assignments
* Make arrangements to complete missing assignments with the original instructor by a mutually agreed upon date but within one calendar year

Completion of a CLAS Course Completion Agreement is strongly suggested. Incompletes cannot be awarded that stipulate: (1) a student may repeat the entire course, (2) repeat or replace existing grades, (3) allow the student an indeterminate period of time to complete a course, or (4) allow the student to repeat the course with a different instructor. The CLAS Course Completion Agreement is available from the CLAS Advising Office, NC 2024.

**Title IX**
T​he University of Colorado Denver is committed to maintaining a positive learning, working and living environment. The University prohibits discrimination on the basis of race, color, national origin, sex, age, disability, pregnancy, creed, religion, sexual orientation, veteran status, gender identity, gender expression, political philosophy or political affiliation in admission and access to, and treatment and employment in, its educational programs and activities, working, and living environment. University policy prohibits sexual misconduct, including harassment, domestic and dating violence, sexual assault, stalking, or related retaliation. If you have experienced some sort of sexual misconduct or discrimination, please visit the Office of Equity/Title IX web site to understand the resources available to you or contact the Office of Equity/Title IX Coordinator (303-315-2567).

**Student Code of Conduct**: Students are expected to know, understand, and comply with the ethical standards of the university, including rules against plagiarism, cheating, fabrication and falsification, multiple submissions, misuse of academic materials, and complicity in academic dishonesty. For suggestions on ways to avoid academic dishonesty, please see the Academic Honesty Handbook: [http://www.ucdenver.edu/faculty\_staff/faculty/center-for-faculty- development/Documents/academic\_honesty.pdf](http://spdev.ucdenver.edu/faculty_staff/faculty/center-for-faculty-development/Documents/academic_honesty.pdf)

* Briefly, plagiarism is the use of another person’s words or ideas without crediting that person.
* Evidence of cheating or plagiarism on any assignment will result in an “F” for the course.

**UC Denver History Department Policy on Plagiarism:** The history department assumes that all work is your own. Any words or ideas that you take from another source, including lectures or outside assistance, need to be acknowledged or cited. Not doing this is considered plagiarism. Plagiarism includes absence of attribution when quoting sources; rewording another person’s ideas and implying they are your own; utilizing the argument or structure of a text without citation; attempting to conceal the degree to which a source has been used; obtaining work from a website and submitting it as your own; and submitting the work of another student, with or without that student’s permission. Plagiarists may face failure of the specific assignment, failure of the entire course, and referral for University discipline, which may result in suspension or expulsion. As these rules merely reflect common sense and ethics, ignorance cannot mitigate the plagiarist’s culpability.

For further clarification of the issue see *The CLAS Academic Integrity and Honor Code:*<http://www.ucdenver.edu/academics/colleges/CLAS/faculty-staff/policies/Pages/AcademicIntegrity.aspx>