



# **You Can't Grow Carrots without Seeds**

**Developing and delivering an open, no-code  
data literacy course for undergraduates**

**Clayton Hayes**

Kenyon College

LOEX Annual Conference, 4 May 2024



# The Soil

-- *Institutional Context*

# Institutional context

## Kenyon



Kenyon College: Small, private,  
liberal arts institution in central  
Ohio

1,885 students, all  
undergraduates

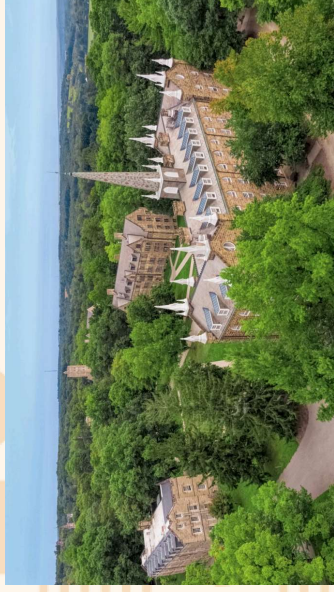
218 FT faculty

11 librarians, 1 archivist

# Institutional context

## Library and Information Services

Kenyon



LBIS division:

- IT
  - Institutional Research
  - Project Management
  - Library
- └─ Instruction & Engagement
- └─ Me



# My context

Prev @ Wayne State: science liaison, digital publishing librarian

Kenyon's Data Services Librarian since August 2022

"Inherited" predecessor's work on a Data Literacy course proposal



# Library courses @ Kenyon

*(pre-me)*

Assoc. VP for Libraries serving on Curricular Policy Committee (CPC) since 2013

Librarians teaching credit-bearing courses since 2016

Two existing courses: Critical Info Lit and Open Science

# Institutional scan

Inherited approach: "stats/data software basics + critical thinking"

But what to use?





# Existing approaches

Investigated CS teaching tools like freeCodeCamp & Codecademy

Investigated alt approaches like Data Carpentry & Programming Historian

Struggled to envision overall learning objectives or coherent *literacy* course



I needed to take a step back and ask:

**What is data literacy?**

# Finding the root

-- *Researching data literacy*



# What is data literacy?

"Strategies and best practices for data literacy education" by Risdale et al., 2015

"Incorporating data literacy into information literacy programs" by Prado & Marzal, 2013

"Data information literacy and undergraduates" by Shorish, 2015

"Designing tools and activities for data literacy learners" by Bhargava & D'Ignazio, 2015

# What is data literacy?

Bhargava and D'Ignazio got me thinking more critically about data literacy:

*"...data literacy has been relegated to a set of technical skills, such as reading charts and making graphs, rather than connecting those skills to broader concepts of citizenship and empowerment."*

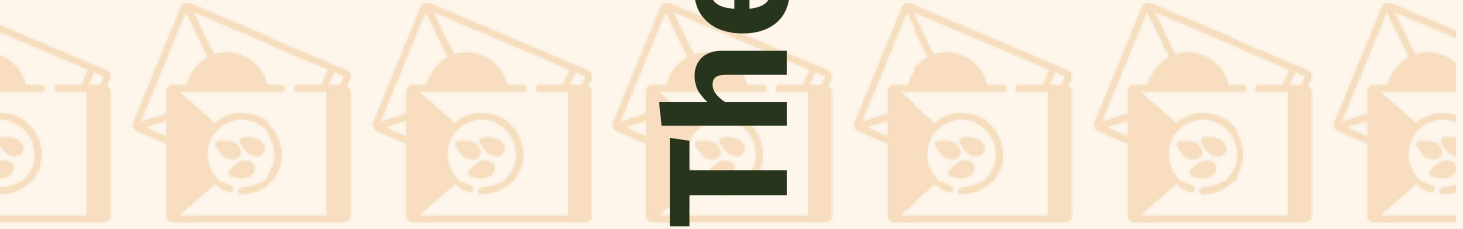
D'Ignazio 2017, p. 7

# What is data literacy?

*"... data literacy includes the ability to read, work with, analyze and argue with data as part of a larger inquiry process."*

D'Ignazio & Bhargava 2016, p. 84

Development of tools and activities supporting data literacy with roots in pedagogical principles



# The seeds of a course

# Data literacy & pedagogy

Existing tools/activities "separate the learning process from the tool itself."

There is a "focus on outputs (spreadsheets, visualizations, etc.), and not on helping novices learn."

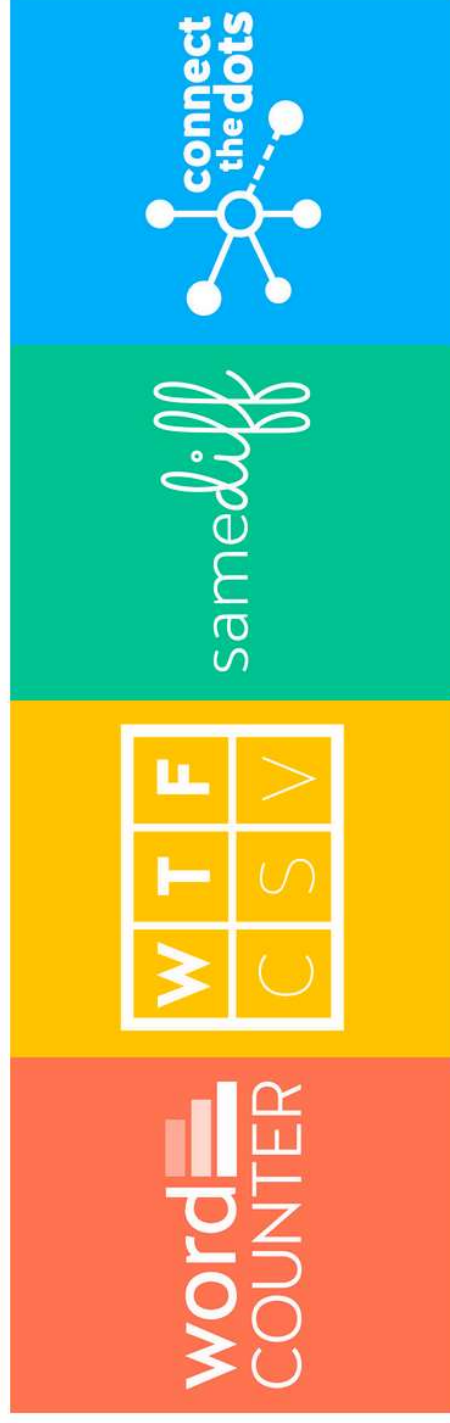
The focus on visualizations prioritizes the quick creation of strong visuals rather than "tools that scaffold a process for learners."

D'Ignazio & Bhargava 2016, p. 85

# DataBasic.io

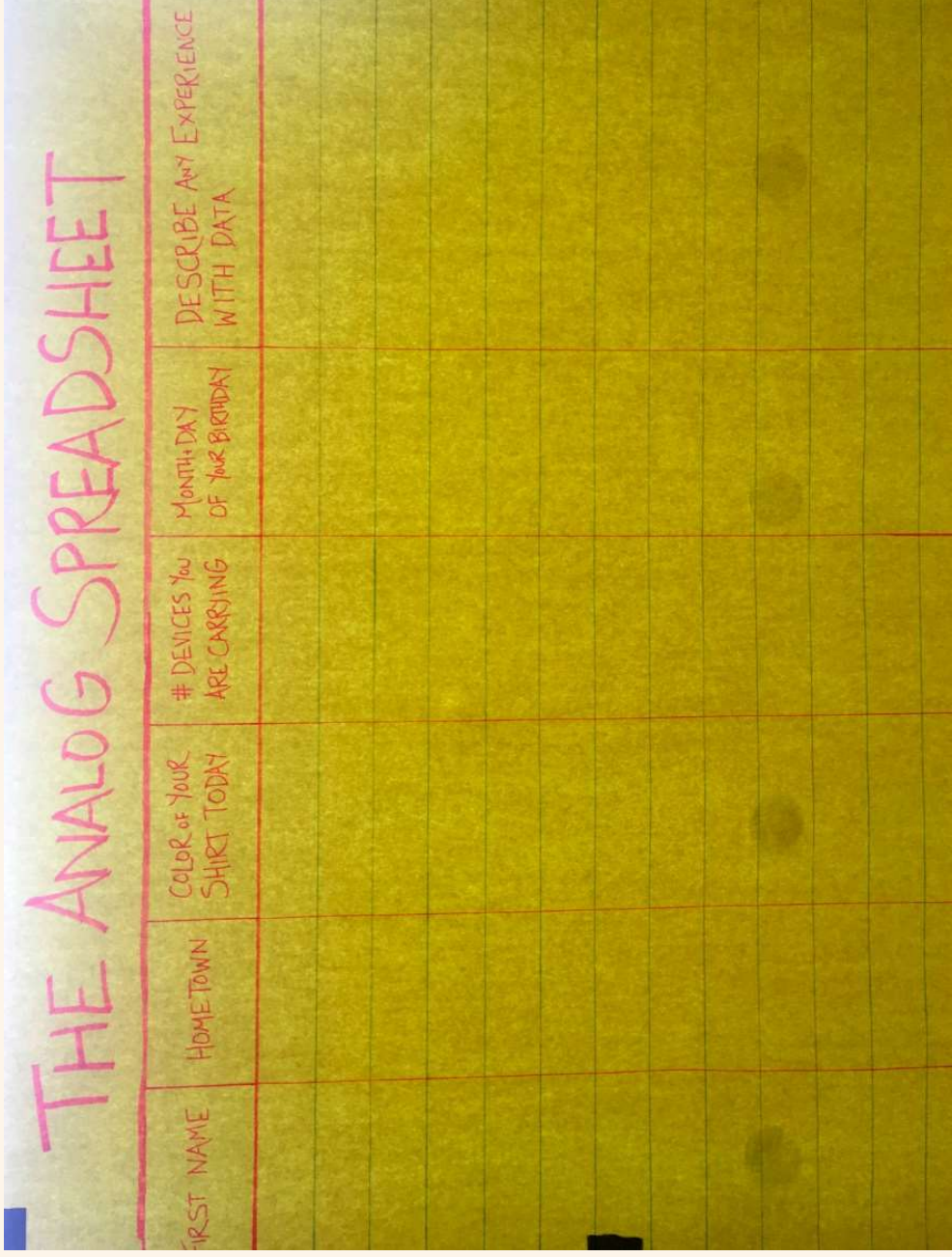
data  
**BASIC.io**

DataBasic is a suite of easy-to-use web tools for beginners that introduce concepts of working with data. These simple tools make it easy to work with data in fun ways, so you can learn how to find great stories to tell.



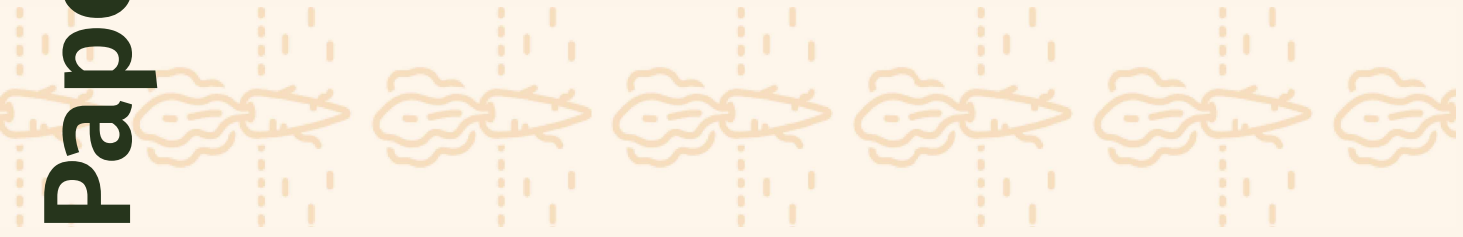


# Paper spreadsheet activity



A hand-drawn spreadsheet on lined paper. The title "THE ANALOG SPREADSHEET" is written in pink at the top. The spreadsheet has six columns and several rows. The columns are labeled with personal information questions. The first row is empty, and the following rows are also empty, providing space for data entry.

FIRST NAME	HOMETOWN	COLOR OF YOUR SHIRT TODAY	# DEVICES YOU ARE CARRYING	MONTH/DAY OF YOUR BIRTHDAY	DESCRIBE ANY EXPERIENCE WITH DATA



# Data literacy & pedagogy

Data murals: Using the arts to build data literacy

*Bhargava et al., 2016*

Feminist data visualization

*D'Ignazio & Klein, 2016*

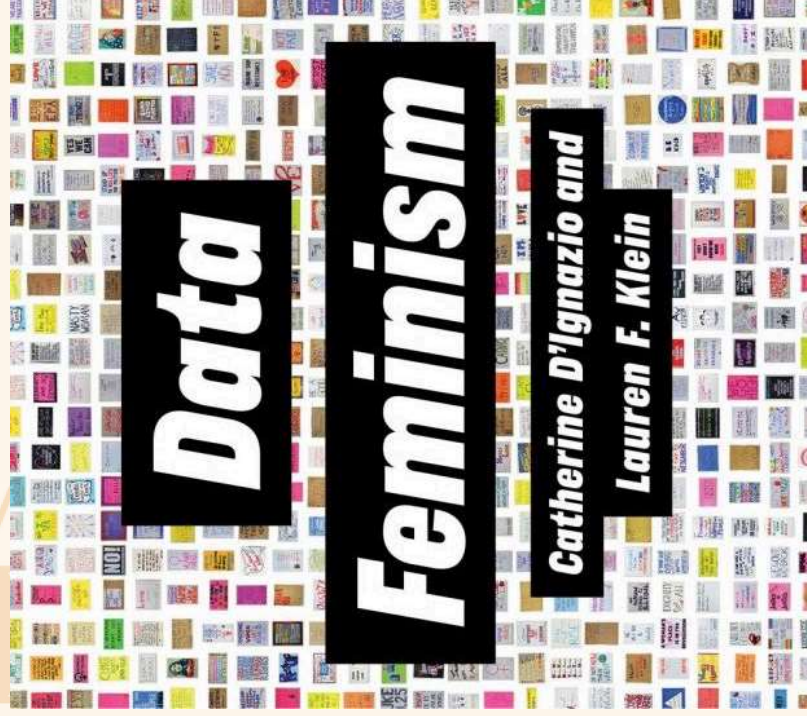
Creative data literacy: Bridging the gap between the data-haves and data-have nots

*D'Ignazio, 2017*

Data visualization literacy: A feminist starting point

*D'Ignazio & Bhargava, 2020*

# The book



*Data Feminism*

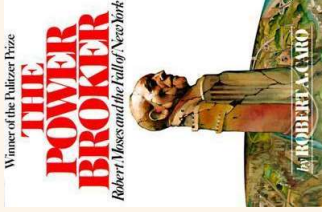
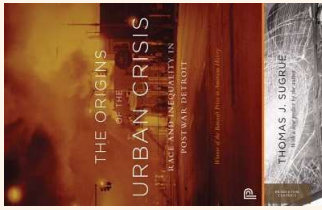
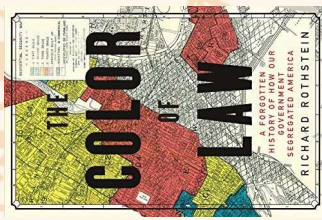
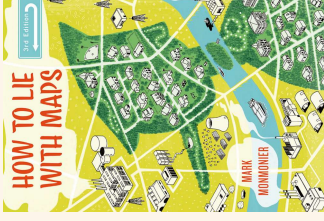
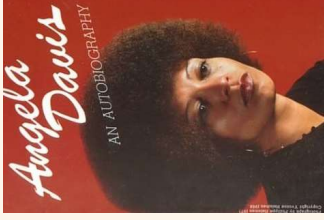
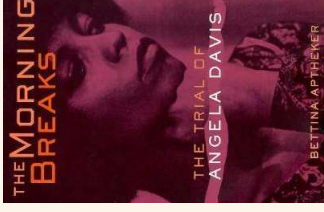
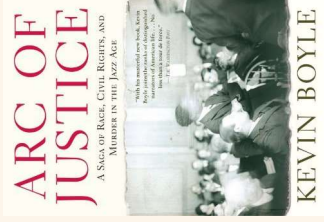
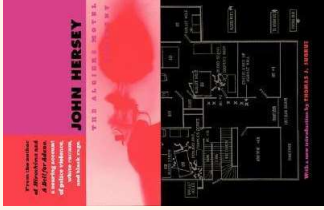
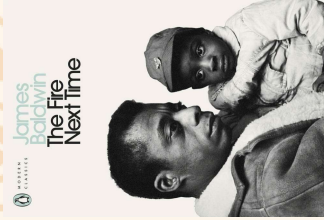
Catherine D'Ignazio &

Lauren F. Klein

2020



# Other books I'd read



# Digging deeper



*All Data are Local* by Yanni Alexander  
Loukissas, 2019

## Data artefacts

Co-presentation by Loukissas and Matthew Battles,  
precursor to *All Data are Local*.



# Deeper!

Feminist data studies: Using digital methods for ethical, reflexive and situated socio-cultural research

*Koen Leurs, 2017*

(Re)framing big data: Activating situated knowledges and a feminist ethics of care in social media research

*Mary Elizabeth Luka & Mélanie Milette, 2018*

# Too deep?

Situated knowledges: The science question in feminism and the privilege of partial perspective

*Donna Haraway, 1988*

"Strong objectivity": A response to the new objectivity question

*Sandra Harding, 1995*

Against cleaning

*Katie Rawson & Trevor Muñoz, 2019*

# **Deciding to move forward**

Research & brainstorming from October to December 2022

By return to work in January 2023 I decided to submit a course

Proposal would be due by 1 March 2023




# The course proposal

*Special Topics* (aka one-off) course proposal consists of:

- Title
- Course description
- Description of assignments
- Weighting of assignments
- Misc. other admin

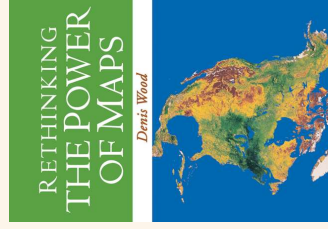
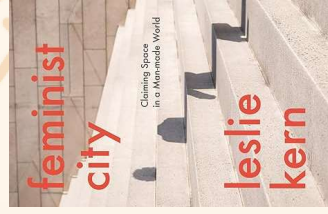
# **INDS 191: Data & Social Justice**

Submitted 13 Feb -- Approved 4 Apr

A decorative border at the bottom of the page features a repeating pattern of stylized orange flowers and green leaves on thin, curved stems.

# Raw & Data & Carrots

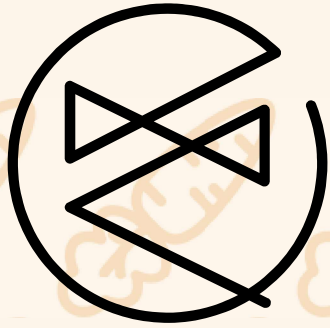
# Even more research!



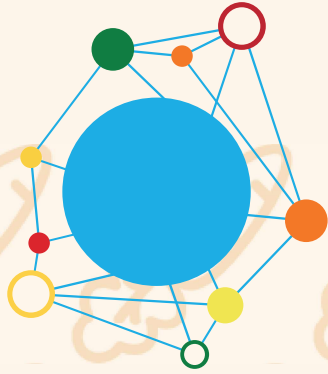
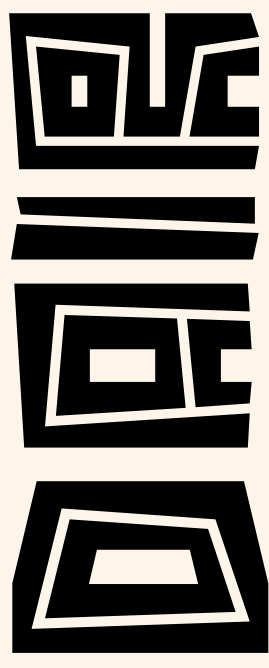
- "Many analysts, one data set"  
*Silberzahn et al., 2018*
- "Extra-activism: Counter-mapping and data justice"  
*Kidd, 2019*
- "On the dangers of stochastic parrots"  
*Bender et al., 2019*
- "Blind spot: Information visualization and art history"  
*Drucker, 2020*

# Even more research!

Organizations to keep an eye on:



ALLIED  
MEDIA  
PROJECTS



DATA-POP  
ALLIANCE

# Even more research!

- "Problem formulation and fairness"  
*Passi & Barocas, 2019*
- "Studying up machine learning data: Why talk about bias when we mean power?"  
*Miceli et al., 2022*
- "Data epistemologies, the coloniality of power, and resistance"  
*Ricaurte, 2019*



*Critical Visualization* by Hall & Dávila, 2023

# Notes from first run

# The students

7 registered:

- 3 seniors
- 2 juniors
- 2 sophomores
- 3 United States:
  - 2 East Coast
  - 1 Ohio
- 4 International:
  - 2 Africa
  - 2 East Asia



# The students

## Majors:

- 2 Art History
- 1 Asian/ME Studies
- 2 Economics
- 2 History
- 1 Internat'l Studies
- 1 Mathematics
- 1 Undeclared

## Minors/concentrations:

- Law & Society
- Mathematics
- Philosophy
- Scientific Computing
- Studio Art

# Student feedback

3 "rounds":

In-class debrief on last day of course

7 (?)

Unofficial google form after last day

3

Official college evaluation at end of semester

3

Not enough structure

More hands-on activities

Covering too much too fast

 Reading on AI

Longer, 2-days/week class

 *Data Feminism*

Course title

# **My thoughts**

Student feedback was (mostly) right

More explicit assignment expectations (especially biography)

Course sequencing was bad

More prep for some activities

Better about "letting go"

# The course repository

<https://github.com/helloitsclayton/data-literacy>



# What's in the box?

- Full syllabus
- Instructor's reading list
- Annotated 8-week lesson plan
- Student reading list
- Ready-to-use data files
- Links to additional materials

# Syllabus

## Social Justice and Data

### Course description

Becoming “data conversant” is only growing in importance as data-related practice becomes increasingly ingrained in public and academic discourse. Learning data-related skills, like reading, working with, analyzing, and arguing with data, is often viewed as a purely technical process that centers mastery of coding or data analysis software.

Instead, this course will take a justice-centered approach that examines how systems of power affect and are affected by data. Students will gain practical experience through low-tech explorations of real-world data, but we won't shy away from or gloss over data's inherent complexities. And data are complex, just like the people who are responsible for their creation, curation, and use.

To help us navigate these murky waters, and to move past number crunching to deeper critical analysis, we'll be drawing on the expertise of scholars, practitioners, and enthusiasts from an array of disciplines and backgrounds.

There are no prerequisites for this course.

### Course goals/objectives

By completing this course, students will have gained:

- Strategies for reading and working with data, as well as experience performing simple data analysis and visualization,
- An understanding of how societal and social factors impact the creation, analysis, and presentation of data,
- Experience leveraging data resources to formulate research inquiries and narratives.

### Texts and readings

# Instructor's reading list

## Where to start

### Important

This is a possible sequence of readings to get you, the instructor, familiar with the topics discussed in the open course "Social Justice and Data." It is a reshuffling of the readings that led to the course's creation, arranged in a way that (hopefully) makes sense.

This is followed by some options for further reading, arranged broadly by category, that I found interesting and/or valuable as I planned the course.

Unless otherwise indicated, all resources are openly accessible.

1. These zines, published by [Allied Media Projects](#), provide accessible introductions to data and data-adjacent topics with a focus on social justice. AMP is a Detroit-based network of individuals and projects and is a leading organization when it comes to design justice.

[Opening Data zine](#) (n.d.), edited by Diana J. Nucera

[Opening Data 2 zine](#) (2017), edited by Diana Nucera and Kristyn Sonnenberg.

2. The *Feminist Data Manifest-No* is a "set of declarations and commitments for feminist data studies," which offers an excellent way in to the principles that underlie this course's approach to data. It also serves as an introduction to a more academic approach to data justice.

[Feminist Data Manifest-No](#) (2019), drafted under the leadership of Marika Cifor and Patricia Garcia.

3. This article introduces the data literacy framework that is at the core of this course, developed by Catherine D'Ignazio and Rahul Bhargava and implemented through their [Data Culture Project](#) collection of resources. (An earlier version of their framework was presented in 2015 at the WebScience Conference in Oxford.)

D'Ignazio & Bhargava (2016) [Databaseic: Design principles, tools and activities for data literacy learners](#) in *The Journal of Community Informatics*

# Student readings

## Week 2

---

### Class 3

- [Children of Immigrants Longitudinal Study](#) by Alejandro Portes & Rubén G. Rumbaut, sections:
  - Summary
  - Scope of Project
  - Methodology
- ["Feeling Unsafe at School"](#) by Cara S. Maffini, sections:
  - Abstract (p. 1)
  - Introduction (pp. 1-2)
  - Participants and Procedures (pp. 5-6)
  - Discussion (pp. 9-11)
- ["Educational Pathway and Social Mobility in Children of Immigrants"](#) by Michael Dunn, sections:
  - Abstract (p. 44)
  - Introduction (p. 44)
  - Sample and Research Design (pp. 45-46)
  - Discussion (pp. 49-50)
  - Conclusion (p. 50)

### Class 4

- ["Yanni Loukissas on understanding and designing data settings."](#) (32:08)  
*Data Materiality* hosted by Scott Rogers ft. Yanni Loukissas
- [Data Feminism Introduction](#), Why Data Science Needs Feminism, pp. 1-19 (18 pages)
- [Data Feminism Chapter 1](#), The Power Chapter, pp. 21-47 (26 pages)



# Data files

The screenshot shows a GitHub repository page for user **helloitsclayton**. The breadcrumb navigation is **data-literacy / students / Class-05**. The commit hash is **cf4aded**, committed 2 weeks ago. The commit message is **rename directory**. The commit history table below lists three files that were renamed in this commit, all 2 weeks ago.

Name	Last commit message	Last commit date
..		
C05-Mapping-Police-Violence-2.csv	rename directory	2 weeks ago
C05-Mapping-Police-Violence.csv	rename directory	2 weeks ago
C05-Ohio-UFO.csv	rename directory	2 weeks ago

# Lesson plan

## Social Justice and Data Lesson Plan

### Important

This lesson plan is structured as an 8-week course, meeting for 90 minutes twice a week.

It was developed as a two-credit course at an institution where 16 credits per semester is the expected course load for students.

- [Activities and Discussions](#)
- [Assignments](#)
- [Materials and Misc. Links](#)

## Week 1

### Class 1 - Course introduction

- Opening activity - "[Paper Spreadsheet](#)" from the Data Culture Project. Have students fill this out as they come into class.
- Go over the course's syllabus and the LMS page for the course.
- Return to "Paper Spreadsheet" for discussion outlined in activity.
- Introduce a few datasets we may be working with regularly?
- At the end of class, assign the "background info" questionnaire via course page, intended to help me learn more about the class' interest/needs, "due" by next class.
- Assigned readings (for next class):
  - [The Myth of Objective Data](#) by Feinberg (2023)
- Assigned readings (to complete by Class 4):
  - [Data Feminism Introduction](#)

# Lesson plan

## Social Justice and Data Lesson Plan

### Important

This lesson plan is structured as an 8-week course, meeting for 90 minutes twice a week.

It was developed as a two-credit course at an institution where 16 credits per semester is the expected course load for students.

- [Activities and Discussions](#)
- [Assignments](#)
- [Materials and Misc. Links](#)

## Week 1

### Class 1 - Course introduction

- Opening activity - "[Paper Spreadsheet](#)" from the Data Culture Project. Have students fill this out as they come into class.
- Go over the course's syllabus and the LMS page for the course.
- Return to "Paper Spreadsheet" for discussion outlined in activity.
- Introduce a few datasets we may be working with regularly?
- At the end of class, assign the "background info" questionnaire via course page, intended to help me learn more about the class' interest/needs, "due" by next class.
- Assigned readings (for next class):
  - [The Myth of Objective Data](#) by Feinberg (2023)
- Assigned readings (to complete by Class 4):
  - [Data Feminism Introduction](#)

# Discussion

## C12 Discussion: Counter-mapping

Maps (and counter-mapping) have already come up in *Data Feminism* a few times: Mariá Salguero's map of femicides in Mexico in Ch. 1 (pp. 37-38), the DGEI and redlining maps in Ch. 2 (pp. 50-51), indigenous place names in Ch. 3 (pp. 93-94), and the Anti-Eviction Mapping Project in Ch. 5 (pp. 127, 129).

1. Why do you think that counter-mapping is such a frequently used means of argument in pushing for social change?
2. What connections can you draw between the counter-mapping projects mentioned (Femicides in Mexico, DGEI, indigenous place names, and the AEMP), the Inuit counter-mapping described in Kidd's article, and Ch. 5's "principle" of data feminism (pp. 125, 147-148)?
3. What about the other principles covered so far? (Examine power, Challenge power, Elevate emotion and embodiment, and Rethink binaries and hierarchies)

### Note

If students need prompting: For Q1, suggest students consider maps as an essential part of the creation of states in the broad political sense; statehood is defined by boundaries on a map, and it makes sense for activists to try and argue using language the state understands. For Q2, you can point out that all of those projects required working collectively, sometimes on a vast scale, and were led by or focused on elevating particular voices.

[Return to Class 12](#)

# Activity (part 1)

## C08 Activity: Planning data cleaning

In this activity, groups will consider how they might go about "cleaning" the data in the Mapping Police Violence dataset, which we looked at in [Class 5](#). It will take groups through common choice points in the cleaning process, and ask them to decide how they might handle each issue.

Provide groups with the following prompt and list of questions:

The first step in most data analysis processes is to "clean" the data so that the process can be more easily automated. For this activity, you'll be looking at the [Mapping Police Violence dataset](#) and considering how you might go about this:

1. Scroll left in the dataset to get a sense for the number of variables (i.e., the number of columns). A description of the different variables can be found in the data documentation. If you were downloading this data set for your own purposes, would you be interested in all of the variables? Which variables might you leave out (and why)?
2. The layout makes it pretty obvious when an individual piece of data is missing, like the "age" of the person in row 9 or the "race" of the person in row 25. What might you do with these gaps? (Also, you can have a different approach depending on the variable/  
column)
3. There are a few different geographic variables in the dataset, some with a lot of variation in how they're recorded (like "street\_address"). If you wanted to associate each data point/incident/row with a single more uniform geographic variable, what might you do?
4. What would you do with the "circumstances" variable/column, which contains plain-text descriptions of the incidents pulled from news reports? Would you keep it as-is, try to reduce it down in some way (maybe come up with categories like for the other variables?), leave the variable out altogether, or something else?
5. BONUS/food for thought: What do you do with data points with more than one thing in them, like "agency\_responsible" in row 14?

[Return to Class 8](#)

# Activity (part 2)

## C07 Activity: "Sketch a Story" with WordCounter

Materials needed:

- Large sheets of white paper
- Poster or other large markers

Though it may not appear so on the surface, the [built-in activity](#) is actually flexible enough to apply to this context. I just (again) had all of the groups work on the same dataset. This also means that comparing how different groups reacted to the same dataset can be a part of the debrief.

- Have students read the brief intro on the [Mapping Inequality](#) site
- Give students some time to explore the [redlining map](#), and be sure to direct their attention to the forms that provide the text they'll be evaluating. This can be done by clicking on a specific area and then switching from the "MAP" tab to the "SCAN" tab.
- For student evaluation: [HOIC Area Descriptions for Ohio](#) from Mapping Inequality

---

### Note

Area descriptions are downloadable on the [American Panorama GitHub](#) via the file `ad_data.json`. I then used OpenRefine to filter to just the entries from Ohio. Because I knew I'd be combined several variables worth of text into a single file (removing the context of column headers) I used OpenRefine to remove entries that just read "None" or something similar. Its "Cluster and edit..." feature is very useful for this.

I exported the resulting data table as a CSV file, and used a spreadsheet editor to copy the text from the following variables into a plain text file: `clarifying_remarks`, `favorable_influences`, `detrimental_influences`. Finally, I did some minimal "cleaning" of the plain text files, mostly just fixing a few character encoding errors.

Usefully, OpenRefine should automatically hang onto your data table; it'll be used in an activity next week.

[Return to Class Z](#)



# Assignments

## Assignment 1: Dataset Biography

A short paper (2-3 pages) introducing a selected dataset and discussing how it came into existence. Please include a link to the dataset - a properly formatted citation in your chosen style would be best - as part of the paper.

As the name implies, you're writing a (concise) biography of that dataset. A good place to start is with the Five Ws: Who, What, When, Where, and Why (and also How). You can follow a similar line of inquiry as we practiced in Class 3 with the CILS data:

1. **Who** - Who was responsible for the creation of the dataset?
2. **What** - What types of data were collected in the study? What variables? What subjects did the study focus on?
3. **When** - When were the data collected?
4. **Where** - Where were the individuals being studied located? Where were the researchers located? Where did data collection take place?
5. **Why** - Why were the data gathered? What is the purpose of the study?
6. **How** - How were individuals selected for the study? How were data gathered?

Be warned, though, many datasets won't just tell you the answers to those questions. In fact, sometimes that information might never have been recorded in the first place. You biography should present what information is available while also commenting on any major gaps you notice.

Below are three very different potential approaches to consider:

- "[Data Biographies: Getting to Know Your Data](#)" by Heather Krause  
A journalistic approach, one that makes sense if you're looking at someone else's dataset
- "[Datasheets for Datasets](#)" by Gebru et al., in *Communications of the ACM*  
Extremely detailed and technical, probably most useful if you generated the dataset
- [Data User Guides from the Western Pennsylvania Regional Data Center](#)  
Also intended for self-generated data, but less technical and more open-ended

*Thank you to...*

---

Dennis Frimpong

Will Margeson

Raya Kenney

Lindsey Neff

Kyumin Kyung

Chau Vu

Behsoy Lowiz



# Image credits

- Kenyon and LBIS logotypes from the [Kenyon Brand Guide](#)
- Ohio map from [Wikimedia Commons](#)
- Old Kenyon residence hall from [Downloadable Images of Kenyon](#)
- Stats/data software logos sourced from [Wikimedia Commons](#)
- Book covers sourced from [Goodreads](#)
- "Carrot" by monkik, from [Noun Project](#); CC BY
- "Carrot" and "field" by Firza Alamsyah, from [Noun Project](#); CC BY
- "Carrot" by Creative Mahira, from [Noun Project](#); CC BY
- "seeds" by ToZIcon, from [Noun Project](#); CC BY
- "Wild carrot flower" by Pham Thi Dieu Linh, from [Noun Project](#); CC BY

# Thank you!

*Full list of cited works (and more!) available  
via course page*

---

Feel free to contact me at:

**hayes3@kenyon.edu**