who we are

who you are

why you were selected

THE CENTER
FOR DIGITAL
HUMANITIES
@PRINCETON

the week's goals

empower humanists to engage with data science concepts

provide hands-on experience in data science methods

help you design a project informed by these methods

THE CENTER
FOR DIGITAL
HUMANITIES
@PRINCETON

9:00 – 10:00	Coffee & Breakfast
10:00 – 12:30	Morning Session
12:30 – 1:30	Lunch
1:30 – 4:00	Afternoon Session

THE CENTER
FOR DIGITAL
HUMANITIES
@PRINCETON

readings

Tuesday: Fundamentals

An introduction to the baseline statistical concepts used in machine learning, from distribution to regression. What does it mean to be "unsupervised"? How do we evaluate a model?

Readings

Meredith Broussard, "Machine Learning: The DL on ML," in *Artificial Unintelligence: How Computers Misunderstand the World* (Cambridge, Massachusetts & London, England: The MIT Press, 2018).

Wendy Hui Kyong Chun, excerpt from *Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition*

THE CENTER
FOR DIGITAL
HUMANITIES
@PRINCETON

Resources

We'll maintain here a running list of resources at the intersection of . . .

Image created by Nick Budak. Inspired by the first issue of Computers & the Humanities, a journal that ran from 1966–2004.

. . . particularly those that focus on 1) the uses of machine learning in humanities research, and 2) critical, historical, and ethical frameworks for analyzing the consequences of machine learning techniques being deployed in the world today.

Contents

- Readings 📖
- Data Sets & Data Work 💾
    - Campus Resources
    - General Portals for Humanities Data
- Technical Instruction ⚙️

Glossary

The following glossary is collaboratively edited by CDH researchers in this Google doc. Because it's still a work in progress, please feel free to comment in the doc to add terms, questions, or clarifications. 🖊️

---

artificial intelligence

> "The field of developing computers and robots that are capable of behaving in ways that both mimic and go beyond human capabilities. AI-enabled programs can analyze and contextualize data to provide information or automatically trigger actions without human interference. Today, artificial intelligence is at the heart of many technologies we use, including smart devices and voice assistants such as Siri on Apple devices." Source

benchmark

> "In this paper we describe a benchmark as a particular combination of a dataset or sets of datasets (at least test data, sometimes also training data), and a metric, concep-

data science

digital humanities

computer vision

data science

natural language processing

data visualization

machine learning

data preservation

the use of **computational methods** in the humanities

digital humanities

&

the **humanistic critique** of emerging technologies

THE CENTER
FOR DIGITAL
HUMANITIES
@PRINCETON

digital humanities

↳ **computational methods**

- automated image recognition using computer vision (Tilton *et al.* 2019)
- NLP: natural language processing (So 2020)
- network analysis (Boswell 2021)
- data visualization, sonification, and physicalization (Hinrichs *et al.* 2018)

THE CENTER
FOR DIGITAL
HUMANITIES
@PRINCETON

digital humanities

↳ **humanistic critique**

- machine learning biases against BIPOC (Benjamin 2019)
- digital surveillance & facial recognition (Boyles 2019)
- technologies of mass incarceration (Ahmed *et al.* 2018)
- misinformation & social media (Hannah 2021)

humanistic critique

- machine learning biases against BIPOC (Benjamin 2019)
- digital surveillance & facial recognition (Boyles 2019)
- technologies of mass incarceration (Ahmed *et al.* 2018)
- misinformation & social media (Hannah 2021)

computational methods

- automated image recognition using computer vision (Tilton *et al.* 2019)
- NLP: natural language processing (So 2020)
- network analysis (Boswell 2021)
- data visualization, sonification, and physicalization (Hinrichs *et al.* 2018)

THE CENTER FOR DIGITAL HUMANITIES @PRINCETON

Of | man's | first diso|bedience | and the | fruit Of |
for|bidden | tree | whose | mortal | taste Brought | de
into the | world | and | all our | woe | Sing | Heavenly | M

Ov mánz férst disobédiens, and ðe frúit 0-1-,½-0,01,00-0-,0-1-
Ov ðát fɵrbíd'n tré, huz mórtal tást 0-½-,01,0-½-,0-1,0-1-
Brét dét intu ðe wúrld, and él ɚr wó ½-1-,00-,0-1-,0-½-,0-1
Wið lós ov 'Ɛd'n, til wun gráter mán 0-½-,0-1,0-0-,0-1,0-1-

" Of | Mān's fīrst | diso|bēdiēnce, | ʃănd thĕ | fru

Δ .

" Of | thāt for|bīddēn | Trēe, whōse | mōrtāl | tāste

Of man's first disobedience, and the fruit
o 2 1 o o 2 o o o 2

Of that forbidden tree, whose mortal taste

author_dob	author_dod	author_fnam	author_lnam	author_role	id	meta_genre	meta_period	meta_rhyme	num_lines	title_edition	title_figure	title_hi	title_main	title_sub
1840	1928	Thomas	Hardy	orig.	Z200136720	Sonnet	Later Ninete	y	14	Collected poems (1930): WESSEX POEN			SHE, TO HIM I	
1840	1928	Thomas	Hardy	orig.	Z200136721	Sonnet	Later Ninete	y	14	Collected poems (1930): WESSEX POEN			SHE, TO HIM II	
1840	1928	Thomas	Hardy	orig.	Z200136722	Sonnet	Later Ninete	y	14	Collected poems (1930): WESSEX POEN			SHE, TO HIM III	
1840	1928	Thomas	Hardy	orig.	Z200136723	Sonnet	Later Ninete	y	14	Collected poems (1930): WESSEX POEN			SHE, TO HIM IV	
1585	1649	William	Drummond	orig.	Z200139958	Sonnet	Tudor 1500-1	y	14	The Poetical Works (1921,Äì1929)			[Like Sophocles (the hearers in a trance)]	
1810	1871	Henry	Alford	orig.	Z200140499	Sonnet	Mid Ninetee	y	14	Poems and Poetical Fragments (1833)			SONNET	WRITTEN AT A DISTANCE FROM HOME.
1810	1871	Henry	Alford	orig.	Z200140502	Sonnet	Mid Ninetee	y	14	Poems and Poetical Fragments (1833)			SONNET.	[Lady, I may not but indite to thee]
1824	1889	William	Allingham	orig.	Z200142302	Sonnet	Mid Ninetee	y	14	Thought and Word [etc.] (1890)			A NURSERY RHYME FOR THE ELDERS.	
1886	1918	Joyce	Kilmer	orig.	Z200142541	Sonnet	Twentieth-C	y	14	[Poems, in] Memories of my son Serge			VALENTINE, 1915	
1794	1878	William Cull	Bryant	orig.	Z200145561	Sonnet	Mid Ninetee	y	14	The Poetical Works (1903)			TO COLE, THE PAINTER, DEPARTING FOR EUROPE.	
1794	1878	William Cull	Bryant	orig.	Z200145718	Sonnet	Mid Ninetee	y	14	The Poetical Works (1903)			A SONNET.	
1779	1843	Washington	Allston	orig.	Z200145979	Sonnet	Early Nintee	y	14	Lectures on art, and poems (1850)			SONNET.	THE FRENCH REVOLUTION.
1779	1843	Washington	Allston	orig.	Z200145980	Sonnet	Early Nintee	y	14	Lectures on art, and poems (1850)			SONNET.	THOUGHT.
1779	1843	Washington	Allston	orig.	Z200145981	Sonnet	Early Nintee	y	14	Lectures on art, and poems (1850)			SONNET.	A SMILE.
1779	1843	Washington	Allston	orig.	Z200145982	Sonnet	Early Nintee	y	14	Lectures on art, and poems (1850)			SONNET.	ART.
1779	1843	Washington	Allston	orig.	Z200145989	Sonnet	Early Nintee	y	14	Lectures on art, and poems (1850)			SONNET	ON THE STATUE OF AN ANGEL, BY BIENAIMvâ,
1779	1843	Washington	Allston	orig.	Z200145990	Sonnet	Early Nintee	y	14	Lectures on art, and poems (1850)			SONNET	ON THE LATE S. T. COLERIDGE.
1779	1843	Washington	Allston	orig.	Z200145991	Sonnet	Early Nintee	y	14	Lectures on art, and poems (1850)			SONNET.	IMMORTALITY.
1779	1843	Washington	Allston	orig.	Z200146018	Sonnet	Early Nintee	y	14	The Sylphs of the Seasons with other p			SONNET On a Falling Group in the Last Judgment of Michael	
1779	1843	Washington	Allston	orig.	Z200146019	Sonnet	Early Nintee	y	14	The Sylphs of the Seasons with other p			SONNET On the Group of the Three Angels before the Tent of	
1779	1843	Washington	Allston	orig.	Z200146020	Sonnet	Early Nintee	y	14	The Sylphs of the Seasons with other p			SONNET On seeing the Picture of vÜolus by Peligrino Tibaldi,	
1779	1843	Washington	Allston	orig.	Z200146021	Sonnet	Early Nintee	y	14	The Sylphs of the Seasons with other p			SONNET On Rembrant; occasioned by his Picture of Jacob's D	
1779	1843	Washington	Allston	orig.	Z200146022	Sonnet	Early Nintee	y	14	The Sylphs of the Seasons with other p			SONNET On the Luxembourg Gallery.	
1779	1843	Washington	Allston	orig.	Z200146023	Sonnet	Early Nintee	y	14	The Sylphs of the Seasons with other p			SONNET To my venerable Friend, the President of the Royal A	
1886	1916	Alan	Seeger	orig.	Z200146054	Sonnet	Twentieth-C	y	14	Poems (1917)			SONNET I	
1886	1916	Alan	Seeger	orig.	Z200146055	Sonnet	Twentieth-C	y	14	Poems (1917)			SONNET II	
1886	1916	Alan	Seeger	orig.	Z200146056	Sonnet	Twentieth-C	y	14	Poems (1917)			SONNET III	
1886	1916	Alan	Seeger	orig.	Z200146057	Sonnet	Twentieth-C	y	14	Poems (1917)			SONNET IV	
1886	1916	Alan	Seeger	orig.	Z200146058	Sonnet	Twentieth-C	y	14	Poems (1917)			SONNET V	
1886	1916	Alan	Seeger	orig.	Z200146059	Sonnet	Twentieth-C	y	14	Poems (1917)			SONNET VI	
1886	1916	Alan	Seeger	orig.	Z200146060	Sonnet	Twentieth-C	y	14	Poems (1917)			SONNET VII	
1886	1916	Alan	Seeger	orig.	Z200146061	Sonnet	Twentieth-C	y	14	Poems (1917)			SONNET VIII	
1886	1916	Alan	Seeger	orig.	Z200146062	Sonnet	Twentieth-C	y	14	Poems (1917)			SONNET IX	

Oh, the humanities!

Share of bachelor's degrees, by major

Humanities
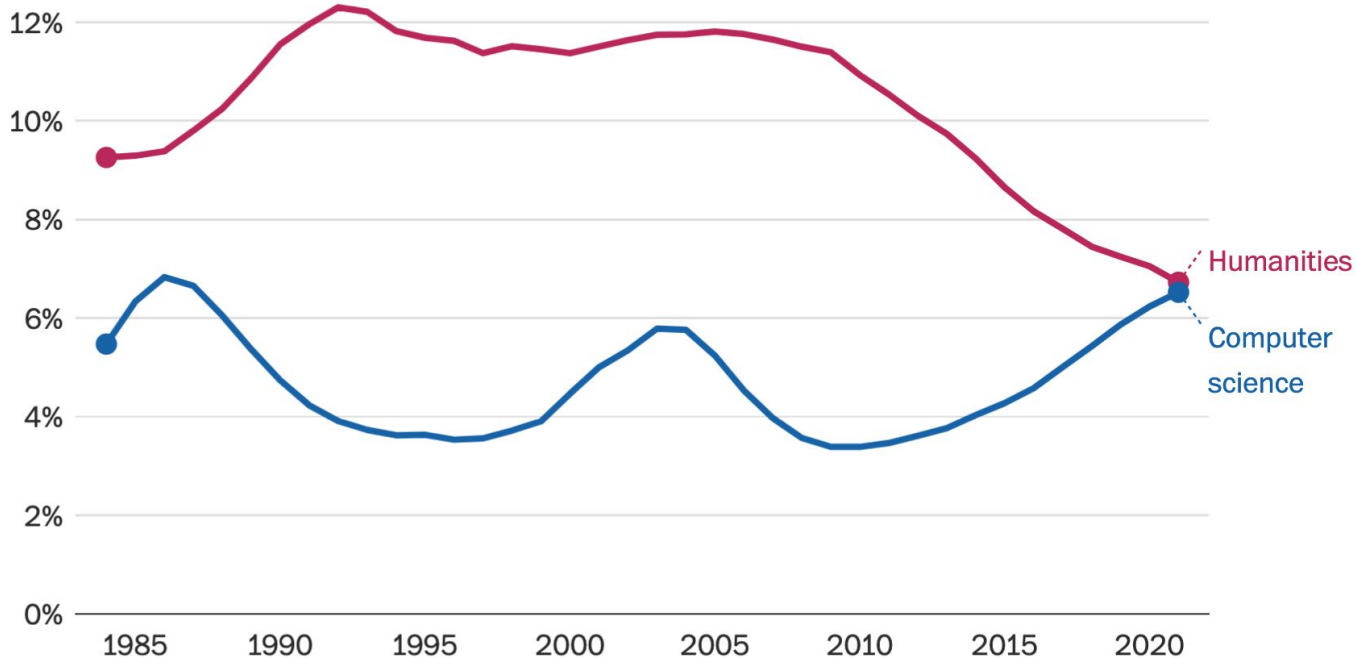
Computer science

Note: doesn't include second majors; computer science includes computer and electrical engineering

Source: Ben Schmidt's analysis of data from the National Center for Education Statistics

DEPARTMENT OF DATA / THE WASHINGTON POST
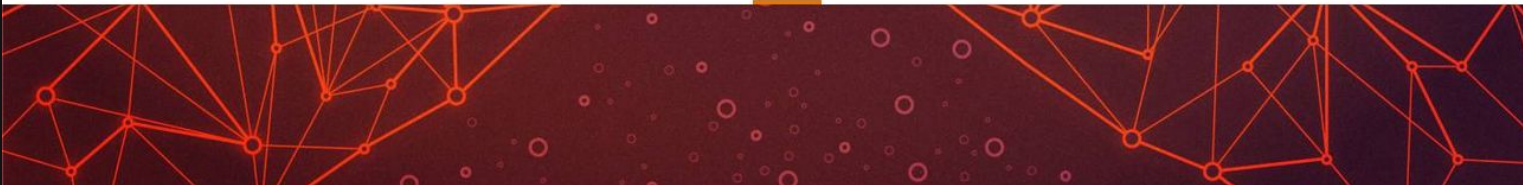
THE CENTER
FOR DIGITAL
HUMANITIES
@PRINCETON

TABLE I.

MILTON	Total No. of words	Per cent Mono- dissyllabic	Per cent Polysyllabic
Paradise Lost I......	5,960	91.67	8.33
II......	7,917	92.24	7.75
III......	5,566	92.07	7.92
IV......	7,700	92.74	7.24
V......	6,804	92.01	7.99
VI......	6,773	90.95	9.03
VII......	4,774	91.40	8.58
VIII......	4,921	91.45	8.53
IX......	9,010	93.01	6.98
X......	8,370	91.74	8.24
XI......	6,859	92.48	7.50
XII......	4,930	91.78	8.21
Total............	79,584	92.03	7.95
Samson Agonistes			
Dialogue............	9,465	92.04	7.94
Choruses............	3,427	90.92	9.08
Total............	12,892	91.75	8.23

Eras in English Poetry

	1500–40	1540–70	1570–00	1600–40	1640–70	1670–00	1700–40	1740–70	1770–00	1800–40	1840–70	1870–00	1900–40
PHRASAL				Fletcher, Sylvester	Milton, More	Wesley, Blackmore	Thomson, Somerville, Dyer, Philips, Prior, Blair	T. Warton, Armstrong, J. Warton, Collins, Lyttleton, Mason, Gray, Akenside	Bowles, Blake	Hemans, Keats	Tennyson, Whitman	Wilde	Moore, Pound
BALANCED	Hawes, Dunbar, Douglas	Sackville	Spenser, Shakespeare, Hall	Sandys	Waller, Crashaw	Dryden, Marvell, Addison, Pomfret	Pope, Parnell, Young	Shenstone, Johnson	Southey, Goldsmith, Crabbe, Wordsworth, Rogers, Burns, Chatterton	Shelley, Bryant, Campbell, Macaulay, Scott	Poe, Arnold, Longfellow, Emerson	Swinburne, Bridges, Hopkins, Phillips, Thompson, Dickinson	E. Sitwell, Yeats, Stevens, Eliot
CLAUSAL	Barclay, Lydgate, Chaucer, Sternhold, Skelton, Ballad, Langland	Lindsay, Gascoigne, Baldwin, Heywood, Wyatt, Surrey, Breton, Googe, Turberville	Sidney, Drayton, Chapman, Marlowe, Daniel, Donne, Jonson	Quarles, Carew, Herrick, Shirley, Harvey, Wither, Herbert	Lovelace, Denham, Vaughan, Cleveland, Suckling, Cowley	Roscommon, Creech, Walsh, Oldham	Gay		Coleridge	Byron, Moore, Landor	Meredith, E. Browning, R. Browning, Rossetti	Hardy, Housman	Cummings, Frost, Auden, Millay

Josephine Miles

Fig 3. From Josephine Miles, "Eras in English Poetry," *PMLA* 70 no. 4 (1955): 861.

DataX effort jumpstarts demonstration data science project at Princeton

Image by Kyle McKernan, Office of Communications

The Office of Communications
Feb. 27, 2019, noon

Princeton University researchers will push the limits of data science by leveraging artificial intelligence and machine learning across the research spectrum in an interdisciplinary pilot project made possible through a major gift from Schmidt Futures.

The Schmidt DataX Fund will help advance the breadth and depth of data science impact on campus, accelerating discovery in three large, interdisciplinary research efforts and creating a suite of opportunities to educate, train

THE CENTER
FOR DIGITAL
HUMANITIES
@PRINCETON

CURRICULUM

With Student Interest Soaring, Berkeley Creates New Data-Sciences Division

By *Alexander C. Kafka* | NOVEMBER 1, 2018

PHOTO BY JUSTIN SULLIVAN, GETTY IMAGES

More and more students at Berkeley are interested in data science, officials there say.

The University of California at Berkeley today announced a new Division of Data Science and Information. It is the university's largest program change in decades and helps secure its status among the country's top data-science research and

CHRONICLE LIST

Recent Private Gifts to Higher Education: U. of Virginia Gets $120 Million for Data Science

By *Chronicle Staff* | JANUARY 27, 2019

Among the latest largest gifts are $120 million to create a new School of Data Science at the University of Virginia, $30 million to help build Spelman College's new Center for Innovation & the Arts, $10 million to provide need-based scholarships at the University of Kentucky, and $6.9 million to support research at the University of Kansas into the earliest human presence in parts of the United States.

THE CENTER
FOR DIGITAL
HUMANITIES
@PRINCETON