Humanities & Data Science Institute

June 5 – 9, 2023

https://princeton-cdh.github.io/hds-institute/
who we are

who you are

why you were selected
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
9:00 – 10:00  Coffee & Breakfast
10:00 – 12:30  Morning Session
12:30 – 1:30  Lunch
1:30 – 4:00  Afternoon Session
= I MIGHT NEED SOME HELP SOON

= SOMETHING IS SERIOUSLY WRONG
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


Resources

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

Computers and the Humanities

*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
digital humanities

the use of computational methods in the humanities
&
the humanistic critique of emerging technologies
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)
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)
Of man's first disobedience, and the fruit
Of that forbidden tree, whose mortal taste
Brought death into the world, and all our woe.

"Of Mān's first disobedience, and the fruit"

"Of that forbidden Tree, whose mortal taste"

Ov mánx férst disobédiens, and àe frúít
Ov dát forbíd’n tré, huz mértal táüst
Brét dét intw ãe wúrlď, and él xar wó
Wid lós ov 'Ed’n, til wun gráter máń

0-1-, ½-0,01,00-0-, 0-1-
0-½-, 01,0-½-, 0-1,0-1-
½-1-, 00-, 0-1-, 0-½-, 0-1-
0-½-, 0-1,0-0-, 0-1,0-1-

"The Center for Digital Humanities @Princeton"
<table>
<thead>
<tr>
<th>author_dod</th>
<th>author_fam</th>
<th>author_lname</th>
<th>author_role</th>
<th>id</th>
<th>meta_genre</th>
<th>meta_period</th>
<th>meta_rhyme</th>
<th>num_lines</th>
<th>title_edition</th>
<th>title_figure</th>
<th>title_hi</th>
<th>title_main</th>
<th>title_sub</th>
</tr>
</thead>
</table>
| 1840       | 1928       | Thomas        | Hardy       | Z200136720| Sonnet     | Later Nineteen | y          |          | 14 Collected poems (1930): WESSEX POE| SHE, TO HIM I
| 1840       | 1928       | Thomas        | Hardy       | Z200136721| Sonnet     | Later Nineteen | y          |          | 14 Collected poems (1930): WESSEX POE| SHE, TO HIM II
| 1840       | 1928       | Thomas        | Hardy       | Z200136722| Sonnet     | Later Nineteen | y          |          | 14 Collected poems (1930): WESSEX POE| SHE, TO HIM III
| 1840       | 1928       | Thomas        | Hardy       | Z200136723| Sonnet     | Later Nineteen | y          |          | 14 Collected poems (1930): WESSEX POE| SHE, TO HIM IV
| 1585       | 1649       | William       | Drummond    | Z200139558| Sonnet     | Tudor 1500- | y          |          | 14 The Poetical Works (1921,Â1929)[| [Like Sophocles (the hearers in a trance)]]
| 1810       | 1871       | Henry         | Alford      | Z200140499| Sonnet     | Mid Nineteen | y          |          | 14 Poems and Poetical Fragments (1833)| SONNET WRITTEN AT A DISTANCE FROM HOME.
| 1810       | 1871       | Henry         | Alford      | Z200140502| Sonnet     | Mid Nineteen | y          |          | 14 Poems and Poetical Fragments (1833)| SONNET. [Lady, I may not but indite to thee]
| 1824       | 1889       | William       | Allingham   | Z200142302| Sonnet     | Mid Nineteen | y          |          | 14 Thought and Word [etc.] (1890) | A NURSERY RHYME FOR THE ELDERS.
| 1886       | 1918       | Joyce         | Klimer      | Z200142541| Sonnet     | Twentieth-C | y          |          | 14 [Poems, in Memories of my son Serge VALENTINE, 1915
| 1794       | 1878       | William       | Culli      | Z200145561| Sonnet     | Mid Nineteen | y          |          | 14 The Poetical Works (1903) | TO COLE, THE PAINTER, DEPARTING FOR EUROPE.
| 1794       | 1878       | William       | Culli      | Z200145718| Sonnet     | Mid Nineteen | y          |          | 14 The Poetical Works (1903) | A SONNET.
| 1779       | 1843       | Washington    | Allston     | Z200145979| Sonnet     | Early Nineteen | y          |          | 14 Lectures on art, and poems (1850)| SONNET. THE FRENCH REVOLUTION.
| 1779       | 1843       | Washington    | Allston     | Z200145980| Sonnet     | Early Nineteen | y          |          | 14 Lectures on art, and poems (1850)| SONNET. THOUGHT.
| 1779       | 1843       | Washington    | Allston     | Z200145981| Sonnet     | Early Nineteen | y          |          | 14 Lectures on art, and poems (1850)| SONNET. A SMILE.
| 1779       | 1843       | Washington    | Allston     | Z200145982| Sonnet     | Early Nineteen | y          |          | 14 Lectures on art, and poems (1850)| SONNET. ART.
| 1779       | 1843       | Washington    | Allston     | Z200145989| Sonnet     | Early Nineteen | y          |          | 14 Lectures on art, and poems (1850)| SONNET ON THE STATUE OF AN ANGEL, BY BIENAIMV,
| 1779       | 1843       | Washington    | Allston     | Z200145990| Sonnet     | Early Nineteen | y          |          | 14 Lectures on art, and poems (1850)| SONNET ON THE LATE S. T. COLERIDGE.
| 1779       | 1843       | Washington    | Allston     | Z200145991| Sonnet     | Early Nineteen | y          |          | 14 Lectures on art, and poems (1850)| SONNET. IMMORALITY.
| 1779       | 1843       | Washington    | Allston     | Z200146018| Sonnet     | Early Nineteen | y          |          | 14 The Sylphs of the Seasons with other pSONNET On a Falling Group in the Last Judgment of Michael
| 1779       | 1843       | Washington    | Allston     | Z200146019| Sonnet     | Early Nineteen | y          |          | 14 The Sylphs of the Seasons with other pSONNET On the Group of the Three Angels before the Tent of Abraham
| 1779       | 1843       | Washington    | Allston     | Z200146020| Sonnet     | Early Nineteen | y          |          | 14 The Sylphs of the Seasons with other pSONNET On seeing the Picture of Vélus by Pergolino Tibaldi,
| 1779       | 1843       | Washington    | Allston     | Z200146021| Sonnet     | Early Nineteen | y          |          | 14 The Sylphs of the Seasons with other pSONNET On Remembrance; occasioned by his Picture of Jacob's D
| 1779       | 1843       | Washington    | Allston     | Z200146022| Sonnet     | Early Nineteen | y          |          | 14 The Sylphs of the Seasons with other pSONNET On the Luxembourg Gallery.
| 1779       | 1843       | Washington    | Allston     | Z200146023| Sonnet     | Early Nineteen | y          |          | 14 The Sylphs of the Seasons with other pSONNET To my venerable Friend, the President of the Royal A
| 1886       | 1916       | Alan          | Seeger      | Z200146054| Sonnet     | Twentieth-C | y          |          | 14 Sonnets (1917) | SONNET I
| 1886       | 1916       | Alan          | Seeger      | Z200146055| Sonnet     | Twentieth-C | y          |          | 14 Poems (1917) | SONNET II
| 1886       | 1916       | Alan          | Seeger      | Z200146056| Sonnet     | Twentieth-C | y          |          | 14 Poems (1917) | SONNET III
| 1886       | 1916       | Alan          | Seeger      | Z200146057| Sonnet     | Twentieth-C | y          |          | 14 Poems (1917) | SONNET IV
| 1886       | 1916       | Alan          | Seeger      | Z200146058| Sonnet     | Twentieth-C | y          |          | 14 Poems (1917) | SONNET V
| 1886       | 1916       | Alan          | Seeger      | Z200146059| Sonnet     | Twentieth-C | y          |          | 14 Poems (1917) | SONNET VI
| 1886       | 1916       | Alan          | Seeger      | Z200146060| Sonnet     | Twentieth-C | y          |          | 14 Poems (1917) | SONNET VII
| 1886       | 1916       | Alan          | Seeger      | Z200146061| Sonnet     | Twentieth-C | y          |          | 14 Poems (1917) | SONNET VIII
| 1886       | 1916       | Alan          | Seeger      | Z200146062| Sonnet     | Twentieth-C | y          |          | 14 Poems (1917) | SONNET IX

THE CENTER FOR DIGITAL HUMANITIES @ PRINCETON
Oh, the humanities!

Share of bachelor's degrees, by major

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

<table>
<thead>
<tr>
<th>1500-40</th>
<th>1540-70</th>
<th>1570-00</th>
<th>1600-70</th>
<th>1640-70</th>
<th>1670-00</th>
<th>1700-40</th>
<th>1740-70</th>
<th>1770-00</th>
<th>1800-40</th>
<th>1840-70</th>
<th>1870-00</th>
<th>1900-40</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PERASAL</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fletcher</td>
<td>Wesley</td>
<td>Blackmore</td>
<td>Blair</td>
<td>Bowles</td>
<td>Blake</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sylvester</td>
<td>More</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hawes</td>
<td>Spenser</td>
<td>Shakespeare</td>
<td>Waller</td>
<td>Dryden</td>
<td>Pope</td>
<td>Shenstone</td>
<td>Southey</td>
<td>Shelley</td>
<td>Poe</td>
<td>Swinburne</td>
<td>E. Sitwell</td>
<td></td>
</tr>
<tr>
<td>Dunbar</td>
<td>%</td>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Douglas</td>
<td>Sackville</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>Barclay</td>
<td>Lindsay</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>Lydgate</td>
<td>Gascoigne</td>
<td>Sidney</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>Chaucer</td>
<td>Baldwin</td>
<td>Drayton</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>Sterndale</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>Skelton</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>Ballad</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>Langland</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
</tbody>
</table>

DataX effort jumpstarts demonstration data science project at Princeton

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

By Alexander C. Kafka | NOVEMBER 1, 2018

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 hopes to secure its status among the country's top data science research and
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.