

The Lexicon of DH

Mary Catherine Kinniburgh

GC Digital Fellow

mckinniburgh@gmail.com

@mckinniburgh

Assisting: Patrick Sweeney

GC Digital Fellow

psweeney@gradcenter.cuny.edu

@pswee

GO TO:

[http://tinyurl.
com/lexiconofdh](http://tinyurl.com/lexiconofdh)

*use firefox as browser for
workshop*



Lexicon

fig. (a) The vocabulary proper to some department of knowledge or sphere of activity; the vocabulary or word-stock of a region, a particular speaker, etc. *(b)* A list of words or names.

~trusty Oxford English Dictionary Online.

DH (Digital Humanities)

Quick search results

No dictionary entries found for 'digital humanities'.

Check your search and try again.

GO

Workshop Goals

We will:

- **consider** what types of project, research, and work are included in the digital humanities,
- **create** a working resource center for our own “lexicon” of digital humanities,
- **explore** a wide range of tools and strategies for research (and project) development,
- **discuss** non-data-driven digital humanities research questions,
- **envision and elaborate** on our own ideas and questions!

Digital Humanities projects use...

digital methods of research

that engage **humanities topics** in their materials

and/or

interpret the results of digital tools from a **humanities** lens

~knowledge is interpretative (Johanna Drucker)

Things to do with DH

- Create community
- Theorize, critique, comment
- Utilize existing platforms
- Develop platforms
- Produce data
- Process data
- Communicate your research in a digital environment
- Collect, curate, and display digital materials

And more...

Collaborative Research

Most of the time, you find out about tools through *people*.

Introduce yourself to the person next to you:

Name

DH Praxis colleague? Or, which program?

Digital skills

Dreaming of what digital projects?

...and then you'll introduce this person to us!



FIRST THINGS FIRST:

ZOTERO

We're covering a lot, and this will make it easy to save URLs, citations, and get your research going.

All these steps work on your own laptop, too.

Work with your partner to make progress together!

Browser Extension

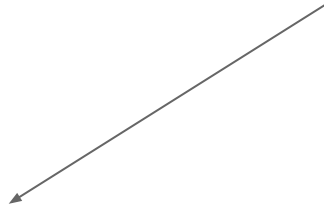


Go to: <https://www.zotero.org/> and click:



Download Now

Available for Mac, Windows, and Linux



Then click:



Zotero for Firefox

Zotero for Firefox lets you capture and organize all your research without ever leaving the browser.



Zotero 4.0 for Firefox

Add a [plugin for Word or LibreOffice](#)

Make Your Account

Success! You Installed Zotero!

New to Zotero? Read the [Quick Start Guide](#) and learn about everything you can do with Zotero.

Register to Take Full Advantage of Zotero

If you haven't already created a Zotero account, please take a few moments to register now. It's a *free* way to [sync and access your library from anywhere](#), and it lets you join [groups](#) and [back up your all your attached files](#).

Your username

Profile URL: <https://www.zotero.org/>

Letter, numbers, spaces

Email

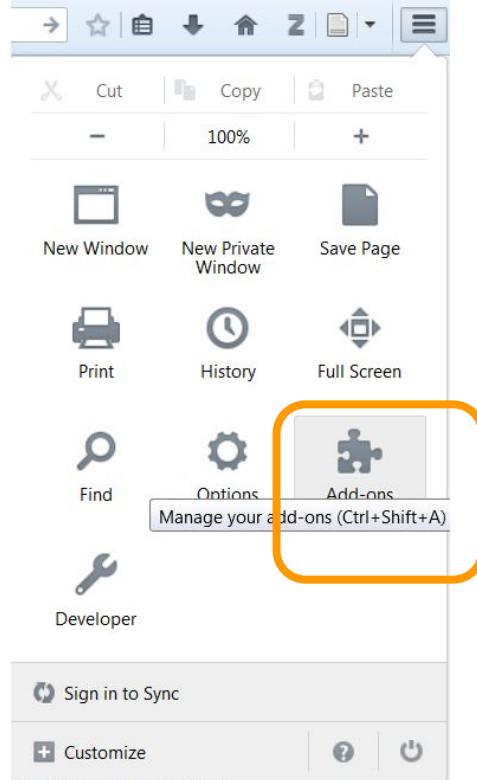
We won't tell anyone

Confirm Email

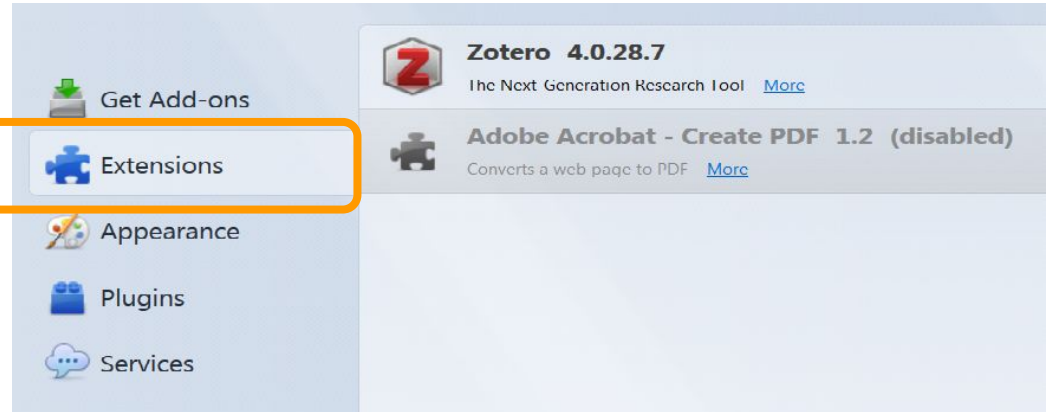
Password

At least 6 characters

Enable the Extension



Go to the upper right hand corner of Firefox to “Add-ons,” then enable Zotero in “Extensions.”



Open Zotero

Once you restart the browser, this will be in the upper right-hand corner.

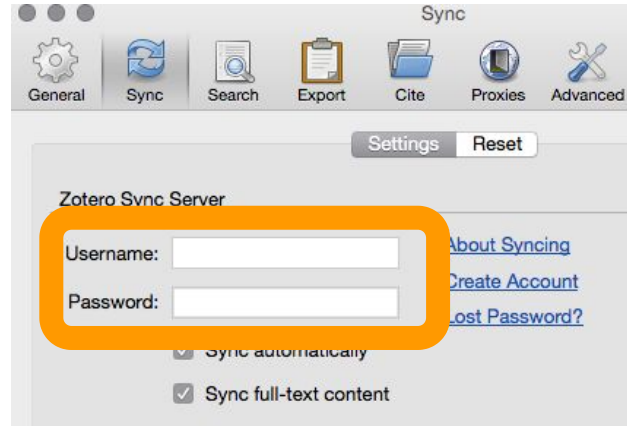
Click on the “Z” button to launch Zotero window. It will appear at the bottom of the browser.



Sync Your Account

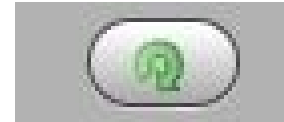


Click on “Preferences” in your Zotero interface near the gear icon.



Fill in the info for the account you just made on Zotero, and close window.

Hit the “refresh” button in the upper right hand corner of the Zotero interface to sync.



Now you're ready to cite! Icons will appear in browser, click to add them to your library.

Okay, ready?

Starting with Data

We'll organize our resource review around the following steps (our grammar):

- **Find** data
- OR **Capture** data

THEN

- **Analyze / Display** data

Assess / Share results.

Messy categories, and we'll discuss why.

Many alternative methodologies, will discuss too.

Make your lexicon:

*Search online for relevant terms as you go, and **add URLs to your Zotero**

***Jot down** phrases, terms, tools, ANYTHING, as you go.

We'll discuss these at the end.

Data for Humanists

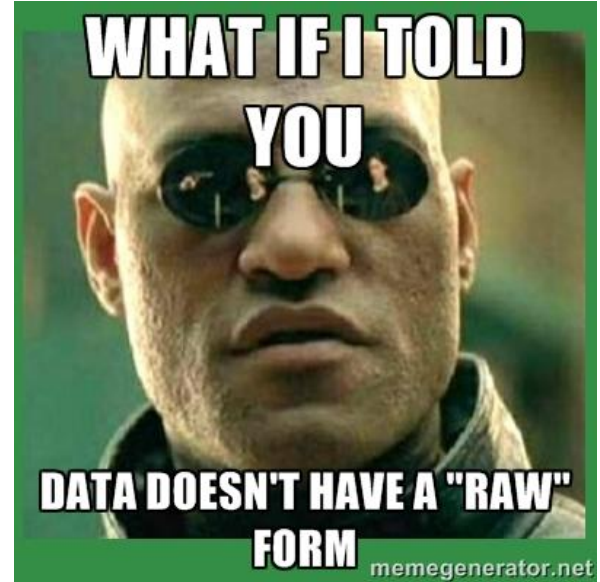
Data is artifact, text, information:

Constructed

Interpretable

Processable (digitally)

Can have evidentiary value



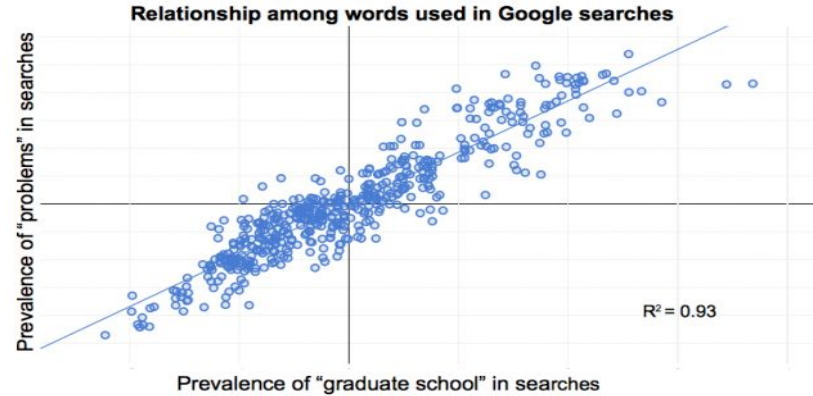
<https://digitalhumanities.stanford.edu/sites/default/files/mor.jpg>

Trevor Owens, "Definine Data for Humanists" *Journal of Digital Humanities* <<http://journalofdigitalhumanities.org/1-1/defining-data-for-humanists-by-trevor-owens/>>

Data in the Wild

Quantitative:

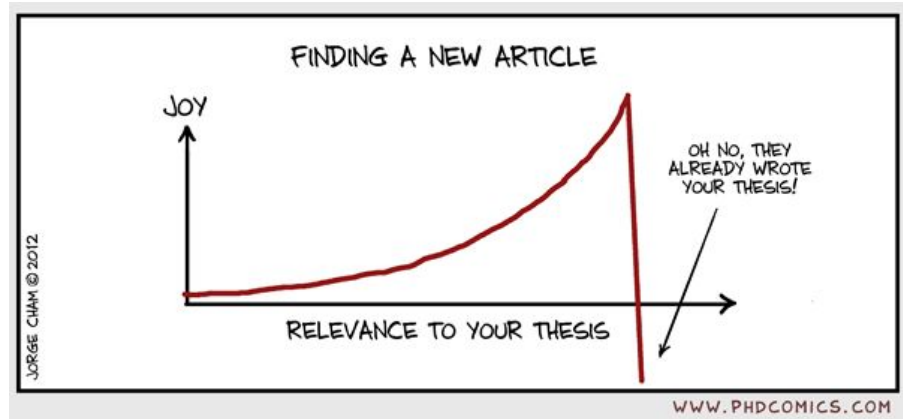
- geospatial points in tabular form
- demographic data
- numerical relations
- image file data



<https://www.insidehighered.com/blogs/gradhacker/mental-health-issues-among-graduate-students>

Qualitative:

- observation
- interviews
- texts
- based on senses



How are you going to get your data?

Find it?

- databases
 - ◆ through libraries
 - ◆ [NYC Open Data](#) (can visualize here too!)
 - ◆ [HathiTrust Digital Datasets](#)
 - ◆ [DPLA \(Digital Public Library of America\)](#)
- institutional repository/collection
 - ◆ [Digital Scriptorium](#)
 - ◆ [NYPL Digital Collections](#)
- other digital projects
 - ◆ [Around DH in 80 Days](#)
 - consider using/examining encoded texts or datasets that other digital projects produce (if open-source or CC-licensed!) to expand research

How are you going to get your data?

Make/Capture it?

- Audio / Visual
- Web Scraping
- Text Analysis
- Text Encoding
- Geocoding

Turn to your partner and quickly brainstorm types of data you might be interested in.

Capturing it with: Audio / Visual

Also: **Take a picture // make a recording // parsing**

uses hardware, software, or even programming language to produce audio or visual content, **OR** uses software or programming language to parse audio or visual content into data.

helps answer questions of: who, what, where, how
to display or share collections, frequency, quantity

Useful URLs:

- [TANDEM](#) (from our own DH Praxis 14/15!)
- [CUNYcast](#) (from our own DH Praxis 14/15!)

Snapshot of Key Words:

→ Scanning

- ◆ to produce your own digital images of analogue content

→ Sound Recording and Analysis

- ◆ [Audacity](#), [Visual Analyzer](#)

→ Image Parsing:

- ◆ The Graduate Center's own: <http://lab.softwarestudies.com/p/software-for-digital-humanities.html>

Capturing it with: Web Scraping

uses programming language to gather specific content from static websites or social media, including image and text.

helps answer questions of: **where (depending), how much, who (depending), frequency**

Useful URLs:

- [Selfie City](#) (used scraping to gather images)
- [No Homophobes](#) (uses scraping to capture homophobic speech on Twitter)
- [Michelle Johnson-McSweeney's recent Digital Fellows workshop and online tutorial](#)

Snapshot of Key Words:

- **API:** Application Program Interface: build
 - ◆ builds software applications
 - ◆ Twitter, Google Maps, Instagram, eBay
- Google sheets offer a way to scrape web and social media (see Michelle's tutorial)
- **Beautiful Soup with lxml:** Python libraries for extracting HTML/XML data
- **Scrapy:** open source framework for data extraction from websites
- **Ethics of Scraping**
 - ◆ may be against site's code of use
 - ◆ consider human element

Capturing it with: Text Analysis

Also: **Topic Modeling // Text Mining**

uses programming languages or software to extract the data, using parameters that you set, from a textual corpus.

*helps answer questions of: **frequency, quantity, style, theme***

Useful URLs:

- [Ted Underwood's Get-Started Guide](#)
- <http://voyant-tools.org/>
- <http://www.tapor.ca/>

Snapshot of Key Words:

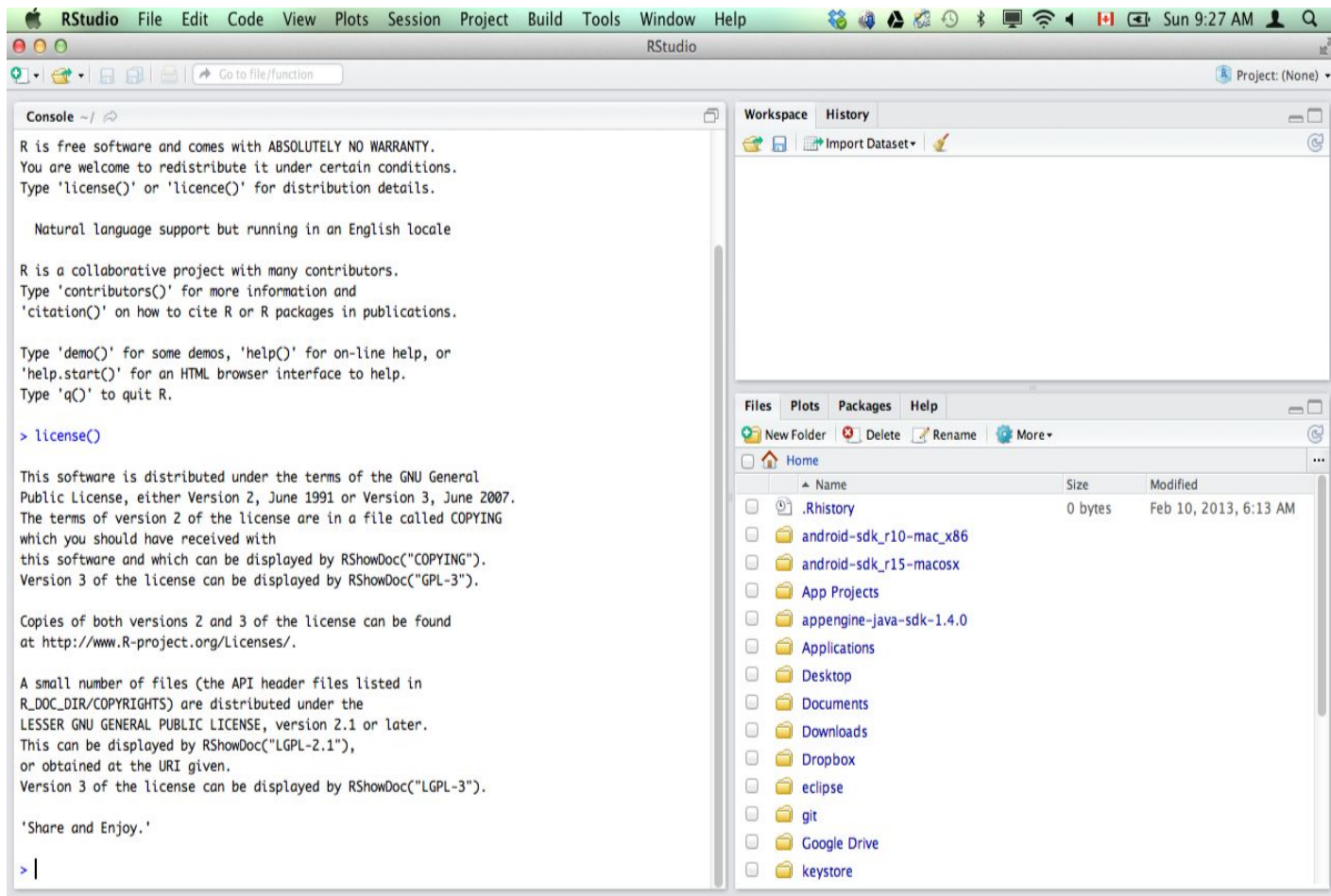
- **AntConc**
 - ◆ corpus analysis and concordancing software
- **R**: statistical computing + graphics environment + language
 - ◆ word frequency, correlation, token distribution
- **MALLET**: Java-based package (cluster of code)
 - ◆ statistical natural language processing, document classification, clustering, topic modeling, information extraction, and other machine learning applications to text.
- **Python**: multi-use programming language
 - ◆ **NLTK (Natural Language Toolkit)** for text analysis

EXAMPLE

R Studio:

GUI (Graphical User Interface)

Command Line: uses text instead of point-and-click



Capturing it with: Text Encoding

Also: **Markup**

uses markup language (usually XML) to encode a text for specific details which often have an interpretive element, or for display in online environments.

helps answer questions of: **whatever you encode for, issues of display**

Useful URLs:

- [Text Encoding Initiative](#)
- [TEI Bibliography of Projects, Theory, and Resources](#)
- [Example: Willa Cather Archive](#)

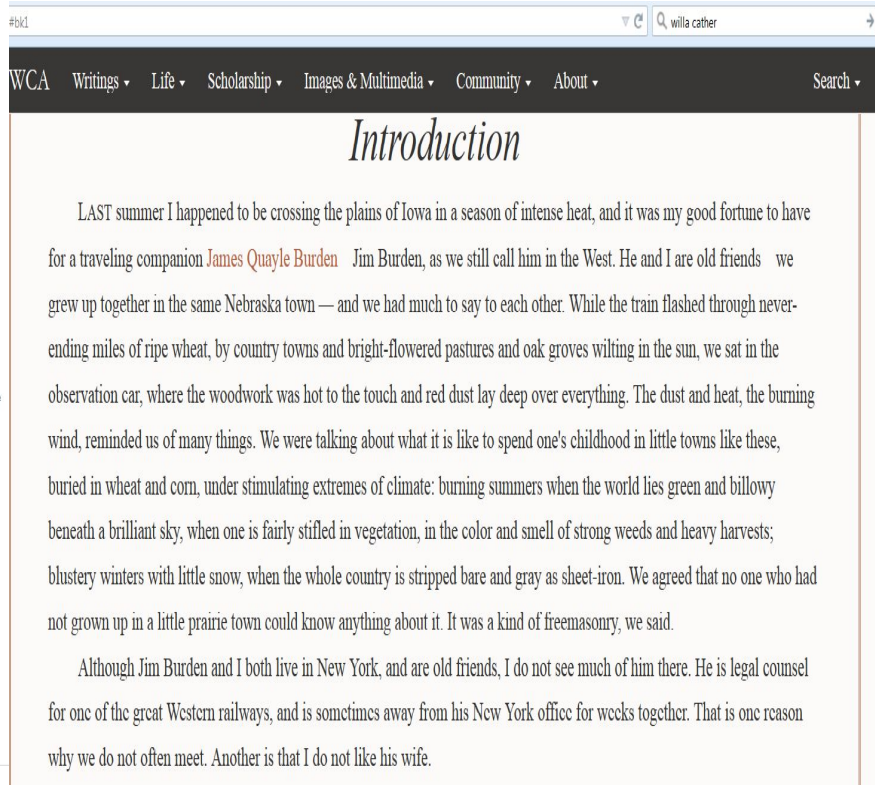
Snapshot of Key Words:

- **TEI:** The Text Encoding Initiative, a standardized system of markup in XML for encoding texts.
 - ◆ archival/museum standard for digital texts
- **XML:** eXtensible Markup Language
 - ◆ HTML's more flexible cousin, which allows you to write your own tags
 - **hence, need for TEI**

Want to Know How It's Made? Right-Click for Page Source

Willa Cather, *My Antonia*

```
463
464
465 </div>
466 <div class="col-md-6 col-centered">
467   <a href="#bk5">
468     <em>Cuzak's Boys</em>
469   </a>
470
471 </div>
472 </div>
473 </div>
474
475 <span class="center"><em>
476   <h1>Introduction</h1></em></span>
477
478 <p><small>&AST</small> summer I happened to be crossing the plains of Iowa
479 in a season of intense heat, and it was my good fortune to have for a traveling
480 companion <a href="#noteen2">James Quayle
481 Burden</a><a class="internalAnchor" name="r2"></a>—Jim Burden, as we still call him in the West. He and I are
482 old friends—we grew up together in the same Nebraska town — and
483 we had much to say to each other. While the train flashed through never-ending miles
484 of ripe wheat, by country towns and bright-flowered pastures and oak groves wilting
485 in the sun, we sat in the observation car, where the woodwork was hot to the touch
486 and red dust lay deep over everything. The dust and heat, the burning wind, reminded
487 us of many things. We were talking about what it is like to spend one's childhood
488 in
489 little towns like these, buried in wheat and corn, under stimulating extremes of
490 climate: burning summers when the world lies green and billowy beneath a brilliant
491 sky, when one is fairly stifled in vegetation, in the color and smell of strong
492 weeds and heavy harvests; blustery winters with little snow, when the whole country
493 is stripped bare and gray as sheet-iron. We agreed that no one who had not grown up
494 in a little prairie town could know anything about it. It was a kind of freemasonry,
495 we said.
496 </p>
497
498 <p>Although Jim Burden and I both live in New York, and are old friends, I do not see
499 much of him there. He is legal counsel for one of the great Western railways, and
```



The screenshot shows a web browser window with the address bar displaying "#bk1" and a search bar containing "willa cather". The website's navigation bar includes links for "WCA", "Writings", "Life", "Scholarship", "Images & Multimedia", "Community", and "About", along with a "Search" button. The main content area features the title "Introduction" in a large, elegant serif font. Below the title, the text begins with "LAST summer I happened to be crossing the plains of Iowa in a season of intense heat, and it was my good fortune to have for a traveling companion James Quayle Burden Jim Burden, as we still call him in the West. He and I are old friends we grew up together in the same Nebraska town — and we had much to say to each other. While the train flashed through never-ending miles of ripe wheat, by country towns and bright-flowered pastures and oak groves wilting in the sun, we sat in the observation car, where the woodwork was hot to the touch and red dust lay deep over everything. The dust and heat, the burning wind, reminded us of many things. We were talking about what it is like to spend one's childhood in little towns like these, buried in wheat and corn, under stimulating extremes of climate: burning summers when the world lies green and billowy beneath a brilliant sky, when one is fairly stifled in vegetation, in the color and smell of strong weeds and heavy harvests; blustery winters with little snow, when the whole country is stripped bare and gray as sheet-iron. We agreed that no one who had not grown up in a little prairie town could know anything about it. It was a kind of freemasonry, we said.

Although Jim Burden and I both live in New York, and are old friends, I do not see much of him there. He is legal counsel for one of the great Western railways, and is sometimes away from his New York office for weeks together. That is one reason why we do not often meet. Another is that I do not like his wife.

Capturing it with: Geocoding

Highlight on Open Source!

Geocoding uses software to transform non-coordinate identifiers (such as landmarks) into geospatial coordinates

[**Geotagging** adds GIS metadata to text, image, or other digital content.]

*helps answer questions of: **where***

Useful URLs:

- [Archipedia \(Google Earth Interface\)](#)
- [Hypercities \(Google Earth Interface\)](#)
- [Saga Map \(QGIS/Leaflet Interface\)](#)



QGIS



free, open-source version of [ArcGIS](#) with many of the same features



OpenStreetMap



an open license, community-sourced mapping service comparable to Google



CartoDB



open source, free option, browser-based



Mapbox



open source, free option



Google Earth / Maps:



API allows for custom applications, can crash when updated

Pause.

Dig Deeper...

Considerations for Capturing:

Take five minutes and

consider a

- database
- method of capturing
- specific tool for capturing
- or hyperlink

from the previous slides or your own web browsing

and address any of the following:

What can you **ask** of this data?

What **can't** you **ask** of this data?

Or, what would it answer **incompletely**?

How **interpretive** is this data / method / tool? How do you know?

For whom would this tool be most useful?

Share your results with your partner and/or the class.

Further Considerations

Do you have the rights to this material?

Do you have a data management plan?

[How are you going to document and standardize your process of collection?]

How does the tool you're using or the method you're using to collect data have the potential to impact your results?

Other considerations?

Analyzing / Displaying Data

Is this a necessary step? Perhaps not always?

*Should we draw a distinction between “**analyze**” and “**display**?”*

~graph / data visualization as argument, demonstration, or image that must be explicated?

Display / Analyze it: Visually

Also: **Data Visualization, Mapping**

Maps dataset as a network, graph, or chart.

Consider the role of your tools in creating your visual argument, as well as your audience.

Popular technique for data journalism.

Further Reading:

Drucker, Johanna. "[Humanities Approaches to Graphical Display](#)." *DHQ*. 5:1 (2011)

Tufte, Edward. *The Visual Display of Quantitative Information*. 2nd Edition. Graphics Press: 2001.

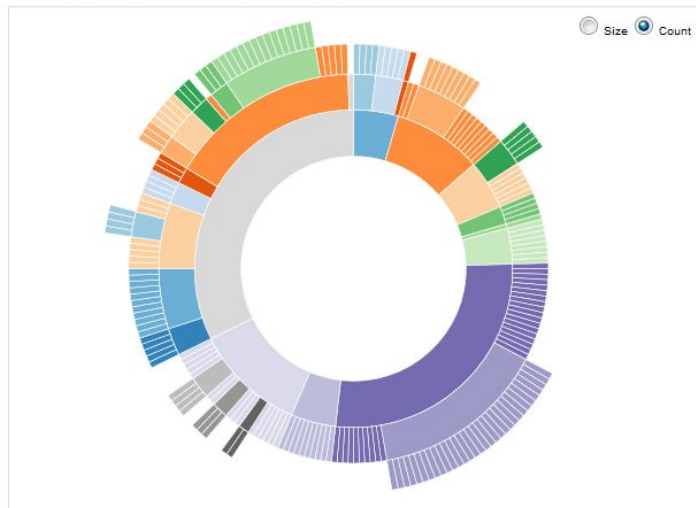
- Excel or other spreadsheets
 - ◆ can generate most basic graphs and charts
 - ◆ upcoming workshop on **Data Visualization** to learn more!
- Word Cloud!
- [Gephi](#)
 - ◆ open-source visualization platform
 - ◆ use **Sigma**, a Javascript library, to render results interactively in a web browser.
- Javascript Libraries
 - ◆ produce visualizations from datasets; numerous options that are open source.
 - ◆ [D3](#)
 - ◆ [dygraphs](#)
- Other tools and platforms on [DiRT Directory](#)

Visualization

Go to <https://github.com/mbostock/d3/wiki/Gallery> and look at images + code to see options

mbostock's block #4083423 November 13, 2012

Sunburst Partition



A sunburst is similar to the treemap, except it uses a radial layout. The root node of the tree is at the center, with leaves on the circumference. The area (or angle, depending on implementation) of each arc corresponds to its value. Sunburst design by John Stasko. Data courtesy Jeff Heer.

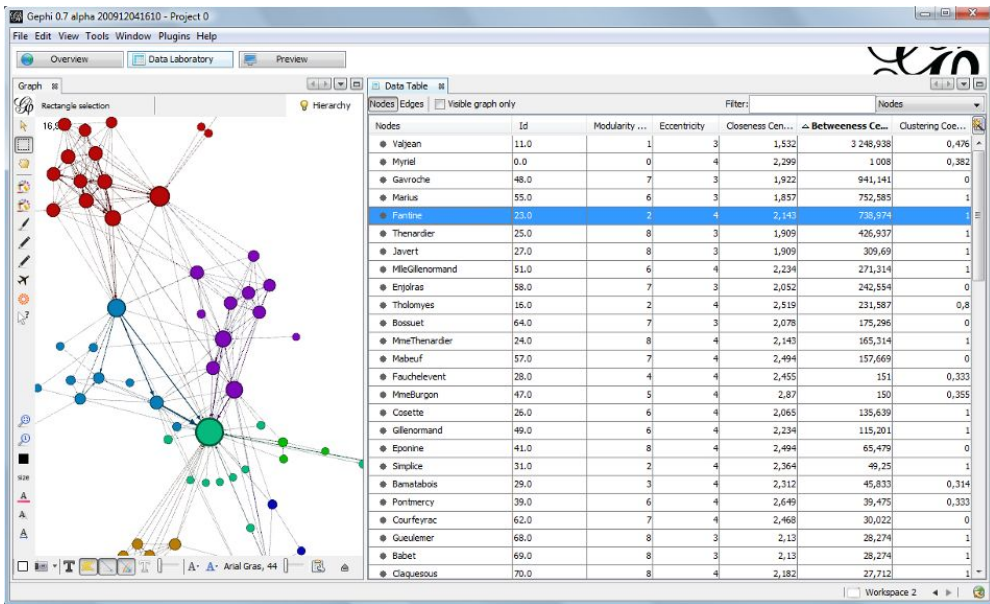
[Open in a new window](#)

index.html

```
<!DOCTYPE html>
<meta charset="utf-8">
<style>

body {
  font-family: "Helvetica Neue", Helvetica, Arial, sans-serif;
  margin: auto;
  position: relative;
  width: 960px;
}

form {
  position: absolute;
  right: 10px;
  top: 10px;
}
```



Inside Gephi (<http://gephi.github.io/images/screenshots/datatable.png>)

Display / Analyze it: Geospatially

Also: **Mapping, cartography, geospatial visualization / analysis**

Renders geospatial coordinates on a projection, with options to change the base map, add color coding, include metadata along with points, and more.

QGIS tutorial on [spatial analysis + map production](#)

Critical GIS: considers processes, assumptions, and history of GIS as well as geospatial data.

[Extensive Bibliography on Critical GIS](#) (and [literary-themed](#)

→ **CartoDB / Mapbox / Google** contain options to render the map in an accessible, browser-based space.

→ **QGIS** requires a Javascript library to render in a browser.

◆ [OpenLayers](#) / [Leaflet](#)

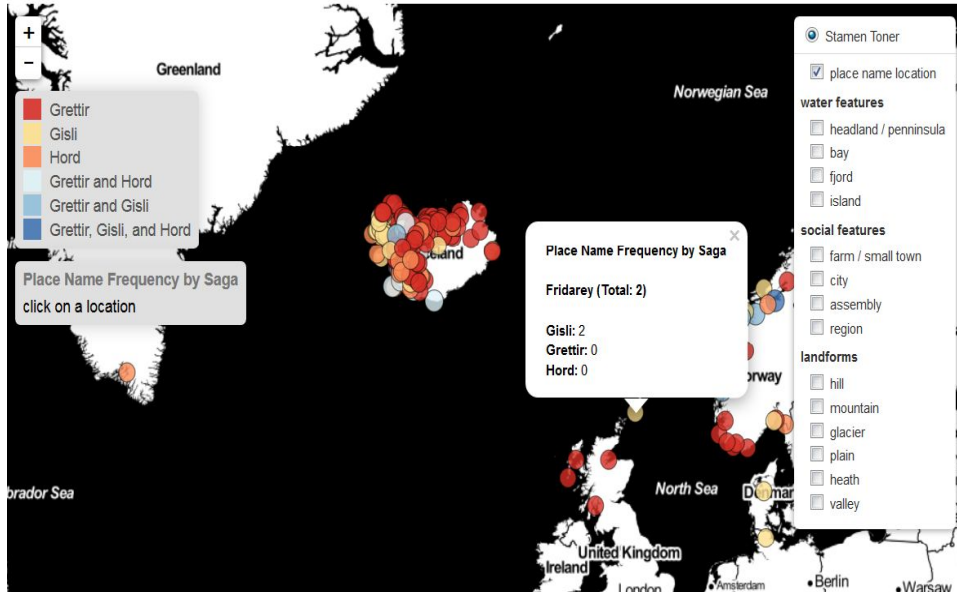
- open source Javascript libraries that use GeoJSON files (produced in QGIS) to display geospatial data in web browsers

→ [Neatline](#)

- ◆ Omeka add-on that allows timelines, maps, and other story-telling features for geospatial information.

Geospatial Maps

Geospatial Mapping to Illustrate Patterns, Metadata



Neatline to overlay maps and archival materials:

["My Dear Little Nelly": Hotchkiss Maps the Battle of Fredericksburg for his Child](#)

Display / Analyze it: Narratively

Also: **Digital Collection, Digital Exhibit, Digital Edition, Web Site, or even essay / article!**

Bring disparate digital (or digitized) materials into a cohesive digital resource.

Enable digital communication of ideas between communities for teaching, collaborating, and learning.

Narrative is one of the most powerful and complex tools we have!

→ Omeka

- ◆ open source platform for exhibitions (library, museum, archives, and scholarly collections)
- ◆ meets Dublin Core metadata standards

→ Scalar

- ◆ multimedia online scholarly publishing platform

→ TEI

- ◆ **XSLT**: eXtensible Stylesheet Language
 - transforms XML (TEI) documents into HTML environments

→ **CMS (Content Management System)**

- ◆ **CUNY Academic Commons**
 - WordPress-based
- ◆ **Drupal**
 - open-source CMS based in **php** (scripting language for web development)
- ◆ [Resources on server architecture / development](#)
from recent GC Digital Fellows Workshop

Pause.

Dig Deeper...

Reflect + Assess

Take five minutes and

consider a

- method of displaying or analyzing
- specific tool
- example of a visualization project
- or any hyperlink

from the previous slides or your own web browsing

and address any of the following:

What does the method of display/analysis **argue**?

How does it **argue**?

What does the method / tool / project make **invisible**?

Is the method / tool **effective**? Why or why not?

Share your results with your partner and/or the class.

Envisioning Other Methods

→ Digital Pedagogy

- ◆ [GC's Interactive Technology and Pedagogy projects page](#)

→ Platform Development

- ◆ [Social Paper](#), [DH Box](#), [CUNY Academic Commons](#), [Commons in a Box](#)

→ Critical theory approaches to tools

→ Assessing current digital tools or datasets

→ Critical Making

- ◆ 3D-Printing, hardware

→ Feminism + DH

→ Game Studies

→ Material Culture Studies

- ◆ [Friedrich Kittler, *Gramophone, Film, Typewriter*. Stanford University Press: 1999.](#)

→ Curating Materials

- ◆ electronic literature, *DHQ* [article by Dene Grigar](#)

Developing Your Lexicon

{determine your topic}

determine data.

{find your data, or make/capture it}

collect data.

{depict this information}

analyze / display data.

{mount your information in a web interface that exists, or develop a platform to either showcase results or allow people to use your same methodology in the same space}

share data.

assess results.

Review your Zotero bibliography, the terms you jotted down, and make a few notes by yourself about projects that you see emerging from these themes.

Discuss with your partner, and share with us!

Ready for More? Resources at the GC

[GC Digital Fellows Office Hours](#) (Wednesdays, 2-3pm in the Digital Scholarship Lab, Room 7414)

Workshops offered by the [GC Digital Fellows](#), [ITP Program](#), the GC [Library](#), and the [Futures Initiative](#)

[GC Digital Initiatives Website](#) and [GC Events and Workshop Calendar on the Commons](#)

More GC-specific and NYC-wide resources available on our [Resource Round-Up Blog](#)

Debrief.

- **developed** a basic “grammar” for DH
- **explored** and **critiqued** methods, tools, and resources
- **discussed** alternative approaches
- **created** a working resource in Zotero

Please take a moment to share your feedback.

<http://tinyurl.com/gcdftopic>

The Lexicon of DH

Mary Catherine Kinniburgh
GC Digital Fellow
mckinniburgh@gmail.com
@mckinniburgh

Assisting: Patrick Sweeney
GC Digital Fellow
psweeney@gradcenter.cuny.edu
@pswee

Survey:
<http://tinyurl.com/gcdftopic>

Slides:
<http://tinyurl.com/lexiconofdh>



THANK YOU!