

The What, Where and How: A Discussion on Digital Humanities and Government Information

Welcome

1. What is Digital Humanities?
2. Examples of Gov Info in Student Projects
3. Tools to Explore
4. Data Resources & Corpora
5. Discussion

Charmaine Henriques, University of Indiana Bloomington

Jesse Silva, University of California Berkeley

Catherine Morse, University of Michigan

Lucia Orlando, University of California Santa Cruz

Adventures in Digital Humanities (from a novice)

- 2009 Literature Grad Student
- 2013 Art History Student
- 2015 Digital Media Undergraduate Thesis Project



The American Presidency Project

ABOUT SEARCH

UC SANTA BARBARA

DOCUMENTS

STATISTICS

MEDIA ARCHIVE

PRESIDENTS

ANALYSES

GIVE

LIBRARY LIBRARY OF CONGRESS

Everything

Library of Congress » Free to Use and Reuse Sets » Free to Use and Reuse: Presidential Portraits

Free to Use and Reuse: Presidential Portraits

Learn more about presidential portraits. Browse more content that is free to use and reuse.

Outreach Tips

- It's ok not to be the digital humanities expert
- Talk with your humanities colleagues
- Talk with your data colleagues
- Check out some of the tools we'll be talking about
- Have fun!

Tools to Explore

- Text Analysis Tools
- Data Visualization Tools
- Digital Mapping Tools

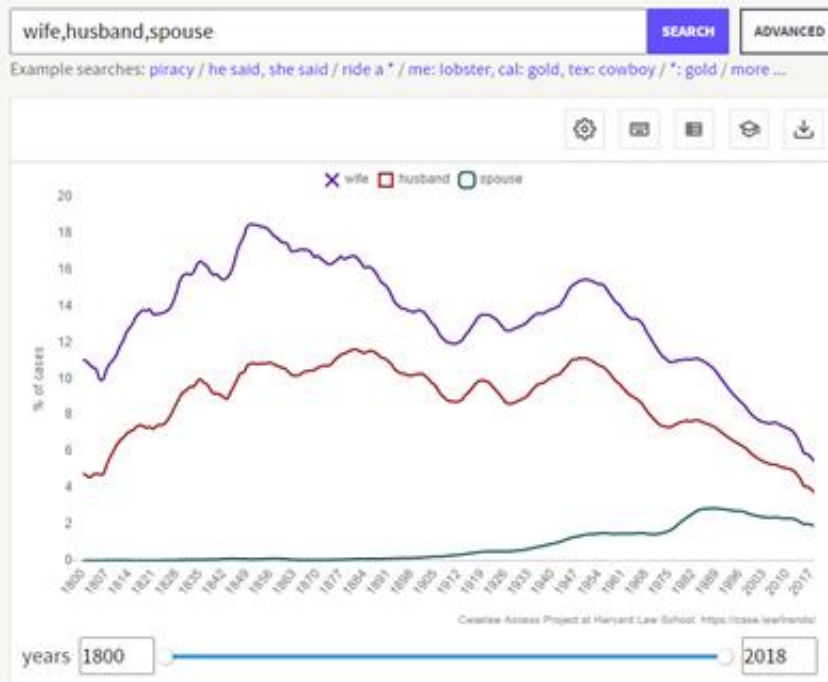
Example: Caselaw Access Project Historical Trends

Similar to Google Ngram Viewer

More at
<https://case.law/trends/>

Historical Trends

The Caselaw Access Project includes over 6 million U.S. legal cases from the Harvard Law School Library — about 12 billion words in all. Our Historical Trends tool graphs the frequency of words and phrases through time, similar to the Google Ngram Viewer.



U.S. Treaties Explorer

Each circle represents a treaty signed by the U.S. Hover over a circle to see more details or click it to visit the treaty document page. To view treaties signed with individual countries, select or type the name of the country in the dropdown menu below. To see bilateral treaties (signed with one other country), type or select **Bilateral** in the menu. To see treaties signed with multiple countries or international organizations, type or select **Multilateral**, **International Organizations**, **Other** in the menu. To see all treaties type or select **All** in the menu.



Example: U.S. Treaties Explorer

Treaty data from Congress.gov → JSON

Used Python to create beeswarm plot

More at <http://gentleknowledge.com/student-projects/congress-treaties/>



Example: Network Analysis

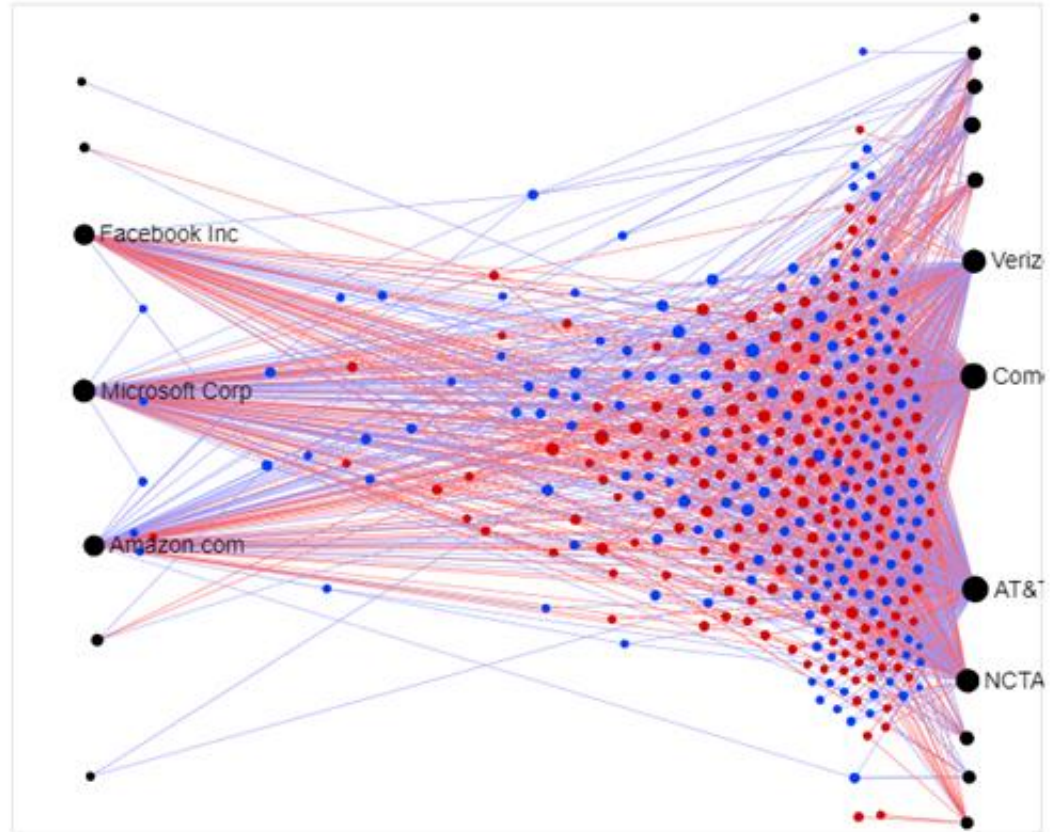
Visualize tabular data

Tools for network analysis: Gephi,
Cytoscape, R

More about this example at:

https://www.opensecrets.org/news/issues/net_neutrality/

Lobbying clients who have lobbied the FCC on net neutrality and their PAC contributions to members of the 115th Congress



Data Visualization Tool

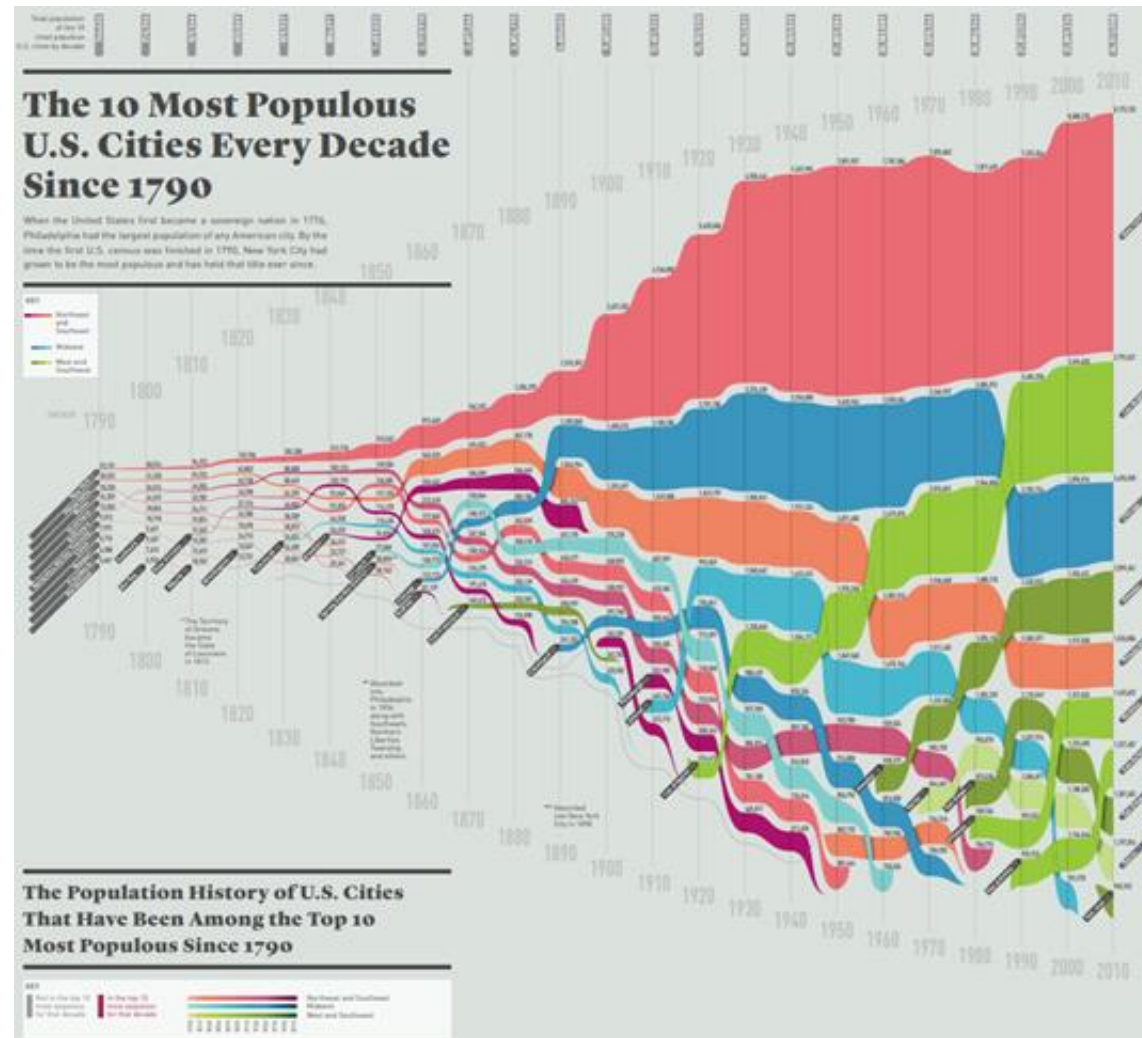
Bring your data RawGraphs

<https://rawgraphs.io/>

Open source data visualization

Make scatterplot, Gantt charts,
bump charts, treemaps etc.

Required formats: tabular data



Make Maps with Data

Tableau Public

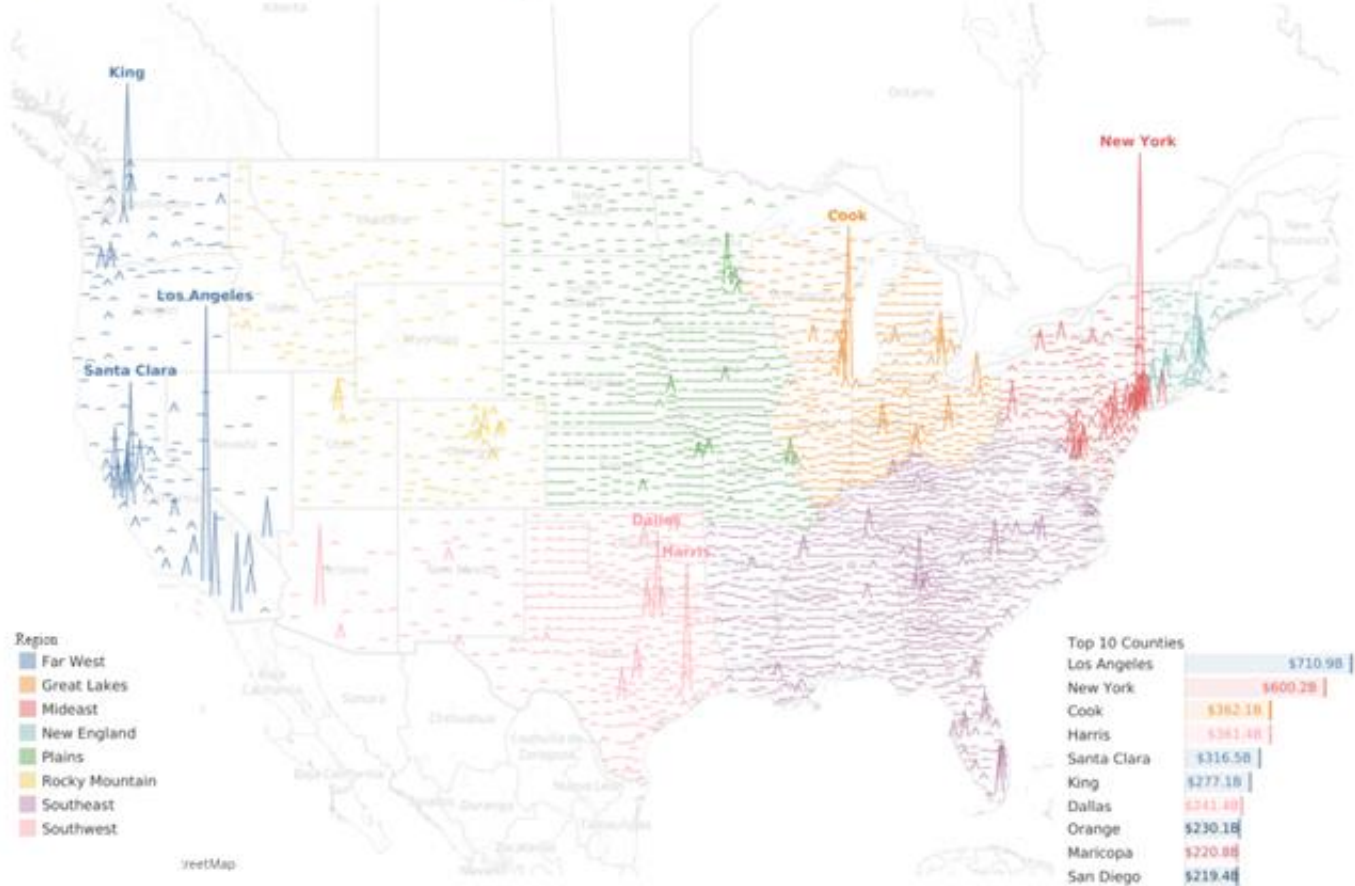
<https://public.tableau.com/en-us/s/>

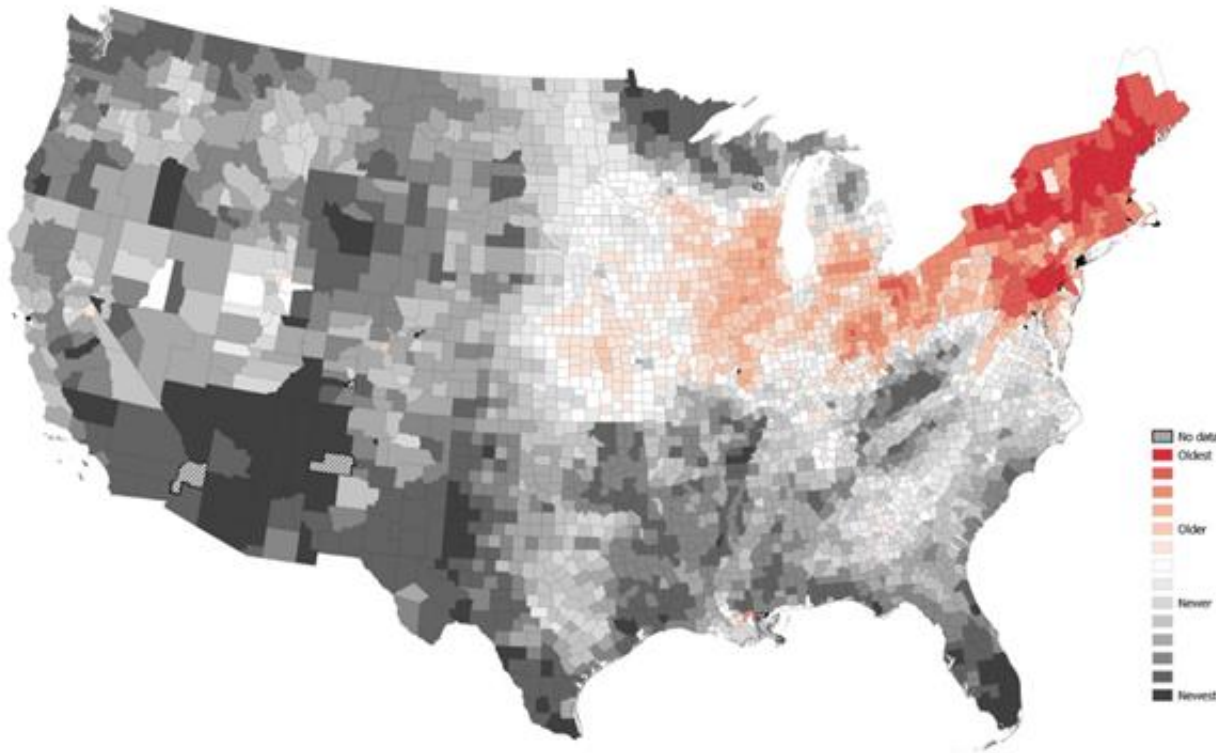
Required Formats:
Microsoft Excel,
multiple text file
formats, statistical
files, Google sheets

Economic output in U.S. Counties

What counties accounted for the most GDP in 2018 across the contiguous U.S.?

Highlight State Highlight County





Make Maps with Data

QGIS

Open source geographic information system

More information at <https://qgis.org/en/site/>

Rural Housing Age in the United States

Data from the 1940 *Census of Housing* and the IPUMS [National Historical Geographic Information System](#). Processed with [QGIS](#).

Gov & Nonprofit Data Sources

Open (Freely Available) Data



- **Chronicling America**
<https://chroniclingamerica.loc.gov/about/api/>
- **GovInfo Bulk Data Repository**
<https://www.govinfo.gov/bulkdata/>
- **American Presidency Project**
<https://www.presidency.ucsb.edu/>
- **GovTrack Data**
<https://www.govtrack.us/about-our-data>
- **HathiTrust Research Center***
<https://www.hathitrust.org/htrc>
- **ICPSR***
<https://www.openicpsr.org/openicpsr/>

And there's more ...

Find More Data

Look for **Libguides** produced by other librarians

Add terms like:

“Open data”

“Digital humanities data”

“Free corpora”

“Textual data”



[UVa](#) / [LibGuides](#) / [Digital Humanities Research](#) / [Finding Tools & Data](#)

Digital Humanities Research

Provides resources for researching topics, tools, and organizations in the Digital Humanities.

<https://guides.lib.virginia.edu/dh/tools>

Berkeley Library

UNIVERSITY OF CALIFORNIA

[Library](#) / [Library Guides](#) / [Government Policy and Information](#) / [U.S. Data and Statistical Sources](#) / [Te](#)

U.S. Data and Statistical Sources: Textual Data

[Search this Guide](#)

Provides access to federal statistics on various subjects and topic areas.

[Introduction](#) [Quick Access to Data/Statistics](#) [Agriculture](#) [Business and Economic](#)
[Criminal Justice](#) [Education](#) [Elections](#) [Environment](#) [Immigration](#) [Labor](#)
[Population and Demographics](#) **[Textual Data](#)**

Free Public Access to Government Information

Free public access to United States government documents in a [federal depository library](#) is guaranteed by law ([44 USC §1911](#)). While UC Berkeley houses one of the most comprehensive collections of U.S.

Sources of Textual Data

This is a starter list to finding textual data from and about the U.S. Government.

- [U.S. Congressional Record \(via Proquest\)](#)
Information on how to access the xml files of the Congressional Record and its predecessor titles. UCB Only
- [FDSys Bulk Data](#)
Various sources of bulk text data from all three branches of government.
- [Congressional and Federal Government Web Harvests](#)
The National Archives and Records Administration

<https://guides.lib.berkeley.edu/US-Stats/text>

Find Data on the Web

Search Google Dataset Search

<https://datasetsearch.research.google.com/>

Suggested Terms:

“Open data”

“Data portal”

... add a geographic area if possible

Ex: California State open data

Or GitHub

<https://github.com/aw-esomadata/>

[aw-esome-public-datasets](https://github.com/aw-esome-public-datasets)

Awesome Public Datasets

NOTICE: This repo is automatically generated by apd-core. Please DO NOT modify this file directly. We have provided a new way to contribute to Awesome Public Datasets. Join the slack community for more communication.

- 🟢 I am well.
- 🔴 Please fix me.

This list of a topic-centric public data sources in high quality. They are collected and tidied from blogs, answers, and user responses. Most of the data sets listed below are free, however, some are not. Other amazingly awesome lists can be found in sindresorhus's awesome list.

Table of Contents

- Agriculture
- Biology
- Climate+Weather
- ComplexNetworks
- ComputerNetworks
- DataChallenges
- EarthScience
- Economics
- Education

Google Dataset Search results for "california state open data".

100+ datasets found

- Legislative Districts in California**
gis.data.ca.gov
Updated Apr 26, 2019
- CA Senate Districts**
gis.data.ca.gov
data-cdegis.opendata.arcgis.com
Updated Dec 13, 2018
- Local Bridges**
gis.data.ca.gov
Updated Jan 15, 2019
- District Tiger Lines**
gis.data.ca.gov
www.wygisctest.org
+1more
Updated Jan 28, 2019
- Traffic Volumes AADT**
gis.data.ca.gov
gis-california.opendata.arcgis.com
Updated Sep 23, 2019

Legislative Districts in California

Explore at gis.data.ca.gov

10 scholarly articles cite this dataset (View in Google Scholar)

Dataset updated Apr 26, 2019

Dataset provided by
[California Department of Education](#)

License
Attribution-NonCommercial 4.0 (CC BY-NC 4.0)
License information was derived automatically.

Area covered

Map Satellite

United States

<https://datasetsearch.research.google.com/>



The first step in text and data mining is to create your Content Set, sometimes known as a corpus or data set. A Content Set is simply a collection of documents you wish to analyze. In this section, we will explore how to use the Gale Digital Scholar Lab to create and curate a Content Set.

1 SEARCH

To start building your content set, you first need to search for documents across the Gale Primary Sources your institution has made available. You can search by keywords or terms that characterize or best describe your research interests, or themes. You can search using words that appear in a text, or the metadata describing it.



Use basic search to quickly find results and employ logic operators like AND, OR, and NOT to further refine your search.



Advanced Search presents you with a framework for building an simple or as complex a search expression as you want. This can search for terms (consisting of one or more words) from one index, or from multiple indexes linked by logical (boolean) operators (AND, OR, NOT). This enables you to retrieve very specific results.

Learn more by reading about Search Strategies

Major vendors also support Text & Data Analysis (TDM) – For a Fee

- JSTOR data for research
<https://www.jstor.org/df/>
 - Free or by Agreement
- Gale Digital Scholarship Lab
- As well as:
 - SpringerLink
 - Proquest
 - Wiley
 - SAGE Journals
 - Taylor & Francis

Communicating with Patrons & Vendors

Text mining is not automatically included with database access - check the license!

Details to Convey to Vendors

- What is the scope and type of information needed?
 - Entire database of news articles? Full run of pdfs from a specific journal?
- How will the data be accessed? Some methods create server issues!
 - Screen scrape? Notify a site administrator first
 - Bulk download? The vendor may need to do it, usually incurs a fee
- One-time occurrence or ongoing need?
 - If ongoing, what frequency? Monthly, annual?
- What is the timeline?
 - Procuring data can take much longer than many researchers realize!

Discussion

Project Examples

Topic Modeling

Obama Word Cloud

<https://obamawhitehouse.archives.gov/blog/2011/01/26/state-union-word-cloud-jobs-america-people-new>

Mapping

Cuban Battlefields

<http://cubanbattlefields.unl.edu/>

Making

3D Model of President Lincoln's Right Hand

<https://www.loc.gov/exhibits/lincoln/journey-to-washington.html>

<https://blogs.loc.gov/thesignal/2020/01/librarians-collections-come-to-life-as-3d-models/>

Word Frequency

Caselaw Access Project Historical Trends

<https://case.law/trends/>

Visualization

U.S. Treaties Explorer

<http://gentleknowledge.com/student-projects/congress-treaties/>

Network Analysis

Open Secrets: Lobbying Clients who have lobbied FCC on net neutrality

https://www.opensecrets.org/news/issues/net_neutrality/

Data Visualization

Raw Graphs

The 10 Most Populous U.S. Cities Every Decade Since 1790, Marco Giannini.

https://rawgraphs.io/gallery_project/the-the-10-most-populous-u-s-cities-every-decade-since-1790/

Mapping

Tableau Public

Economic Output of U.S. Counties

<https://public.tableau.com/en-us/gallery/economic-output-us-counties?tab=viz-of-the-day&type=viz-of-the-day>

Mapping

QGIS

Rural Housing Age in the United States. Christopher Riley. QGIS Map Showcase:

<https://www.flickr.com/photos/60869609@N04/49219293838/in/pool-qgis/>