
Beyond Books: Exploring the Role of Public Libraries in the U.S. Through Data Visualization

Pinyu (Pearl) Hwang
Adviser: Prof. Holly Rushmeier
Computer Science

Abstract

As access to information becomes increasingly digital, the role of public libraries in our society has shifted. Just as the rest of us adapt to advances in technology and trends in society, so have libraries adapted their function and services to better serve their communities, for example increasing the variety in the format of their materials to include digital items such as ebooks, audio, and video. This project aims to examine and present the societal impact of public library systems in the United States through data visualization techniques. The visualizations created in the course of this project are built using HTML, CSS, and JavaScript, in particular the JavaScript library D3.js, as well as Plotly and jQuery. In addition to visualizing statistics specific to public libraries and public library systems such as circulation volume, population served, and size of holdings, this project also explores library statistics in conjunction with census and demographic data such as rate of poverty and education attainment.

Background & Motivations

There aren't many truly public places left in America. Most of our shared spaces require money or a certain social status to access. Malls exist to sell people things. Museums discourage loiterers. Coffee shops expect patrons to purchase a drink or snack if they want to enjoy the premises.

—Jennifer Howard, “[The Complicated Role of the Modern Public Library](#)”

Public libraries form the backdrop of many fond memories for me, and they're one of my favorite places to be. Whenever I have to stay somewhere for longer than a few days, I find

myself either within or looking for its public libraries: New York, NY; Baltimore County, MD; New Haven, CT; Seattle, WA; King County, WA...

And it's not just about the books (and audiobooks, DVDs, even video games), although they are wonderful. The services public libraries provide extend far beyond the multitude of materials they manage and lend out; public libraries provide its patrons with services such as internet and computer access, study spaces, programs, classes, and clubs for people of all age groups and a variety of interests, homework help and online tutoring, college prep services, job-hunting assistance and assistance with other social services, cultural events, access to online resources like databases and streaming services, information about community engagement, etc. The list can go on.

In an [2018 op-ed in the *New York Times*](#), sociologist Eric Klinenberg describes public libraries as *social infrastructure*, “physical places and organizations that shape the way people interact.” Moreover, existing in a time when the explosion of information online can often be confusing and overwhelming, libraries and librarians are still among the top sources of information trusted by Americans, according to [a 2016 study by the Pew Research Center](#).

As such, I am interested in exploring the role that public libraries have played in the U.S., and looking at any trends that may exist. In particular, my visualizations focus on: 1) How has the role of public libraries changed throughout the years? 2) How does public libraries service and use relate to poverty and educational attainment? And 3) How can I design interactive visualizations that are intuitive and informative, and which will allow the user to easily explore and compare statistics of personal interest?

Methods & Implementation

The visualizations in this project are created using HTML, CSS, and JavaScript, in particular the D3.js library; Plotly and jQuery are also used. I also used tools such as Python and Microsoft Excel mainly on a case-by-case basis during much of the data cleaning process.

Sources of Data

The following sources of data were used in this project:

- Institute of Museum and Library Services (IMLS): Public Libraries Survey
<https://www.imls.gov/research-evaluation/data-collection/public-libraries-survey>

- Census.gov: Small Area Income and Poverty Estimates (SAIPE) Program
<https://www.census.gov/programs-surveys/saipe.html>
- Census.gov: American Community Survey (ACS), education attainment data
<https://www.census.gov/programs-surveys/acs>

The main source of federal support for libraries and museums in the U.S., the Institute of Museum and Library Services (IMLS) is an independent agency of the United States federal government established in 1996. The Public Libraries Survey (PLS) has been conducted annually since 1988, first by the National Center for Education Statistics up until 2005, with the IMLS taking over from then on. For my project, I mainly worked with data files from 1998 to 2020, after consideration of variation in the types of data collected over the years. The PLS data comes in sets of three files, with data at the state level, the administrative entity (AE) level (i.e. each public library), and the level of library service outlets. For this project, I mainly worked with data at the state and AE levels.

The Small Area Income and Poverty Estimate (SAIPE) Program produces single-year estimates of income and poverty statistics for school districts, counties, and states. For this project, I used data at the state and county levels for the year 2020.

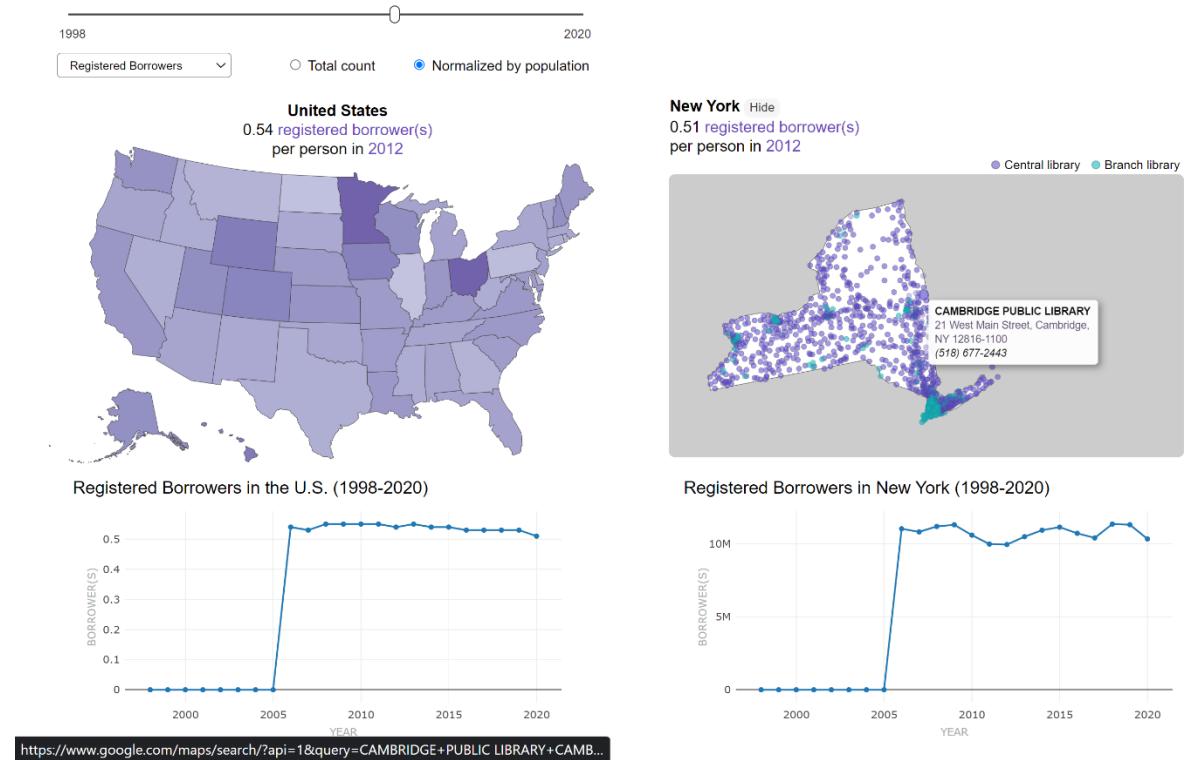
The American Community Survey (ACS) collects population and housing data each year. For this project, I explored education attainment data collected by the ACS at the state and county levels for the year 2020.

Data Extraction

The data collection and extraction process involved identifying data sources, downloading data sources, renaming and reorganizing data files, and cleaning and repackaging data. The most involved aspects of the data cleaning process were twofold. First, it was important that I verify field names in the PLS datasets throughout the years, as some field names change, some fields are split into more specific subfields, and yet other fields are added to the survey. For example, the field ELMATS ‘electronic materials’ was used up until 2002, until EBOOK ‘ebook’ took over in 2003. Second, different datasets had different conventions for county names. This included things such as whether ‘city’ is included in the name, capitalization and spacing (e.g. *De Kalb* versus *DeKalb*), and abbreviations.

Implementation

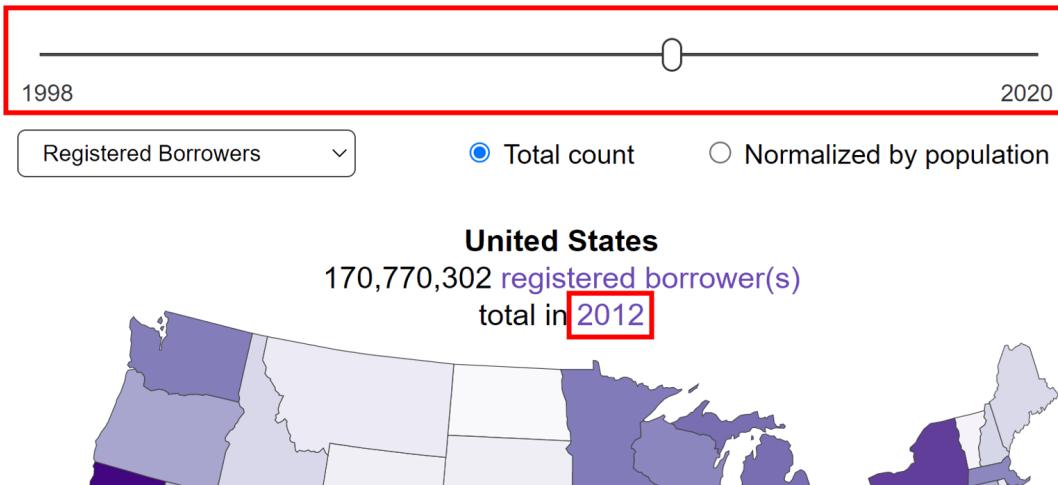
Visualization Set #1



The first set of visualizations explores the PLS datasets from 1998 to 2020 (the most recent year for which a dataset is available as of December 6, 2022). The main idiom for this set of visualizations is a choropleth map, using color to encode various statistics collected through the PLS. Supplementary visualization idioms used are line charts (showing data over the years) and a donut chart (showing the distribution of types of materials). Almost all elements are interactive, including tooltips that display more detailed information on hovered-over or clicked elements. There is a slider for users to select specific years, a dropdown selection tool for choosing individual statistics, and a radio button toggle for total count versus count normalized by population.

Below I outline the ways to interact with this set of visualizations.

Slider

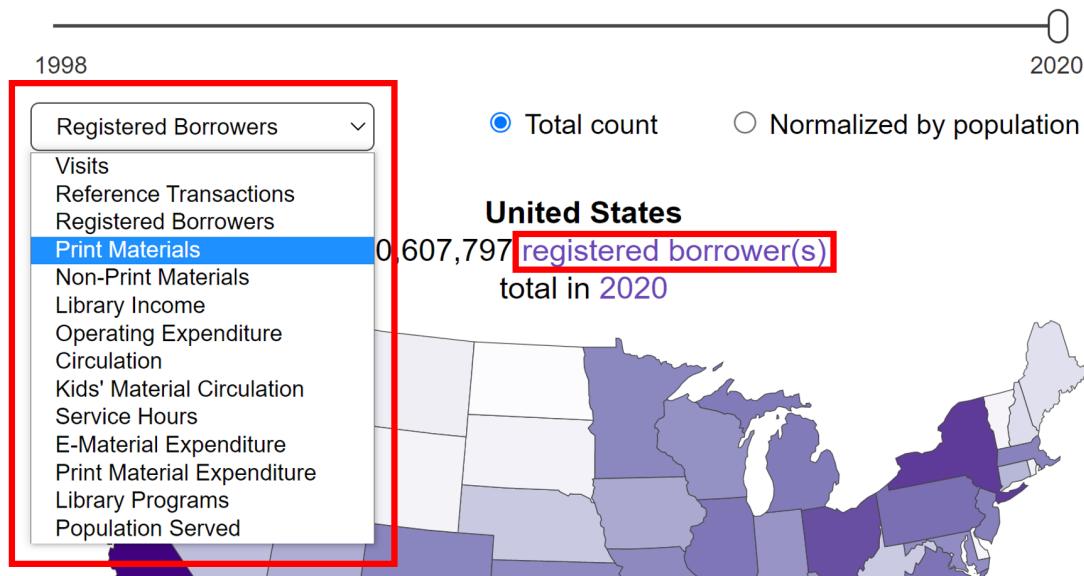


Variable: year

Range of values: 1998-2020

Effect: display data for the selected year

Dropdown



Variable: PLS data field

Range of values:

- VISITS: Visits
- REFERENC: Reference Transactions
- REGBOR: Registered Borrowers

- BKVOL: Print Materials
- OTHMAT: Non-Print Materials
- TOTINCM: Library Income
- TOTOPEXP: Operating Expenditure
- TOTCIR: Circulation
- KIDCIRCL: Kids' Material Circulation
- HRS_OPEN: Service Hours
- ELMATEXP: E-Material Expenditure
- PRMATEXP: Print Material Expenditure
- TOTPRO: Library Programs
- POPU_LSA: Population Served (Population of Legal Service Area)

Effect: configure visualization(s) using selected statistic

Radio Buttons

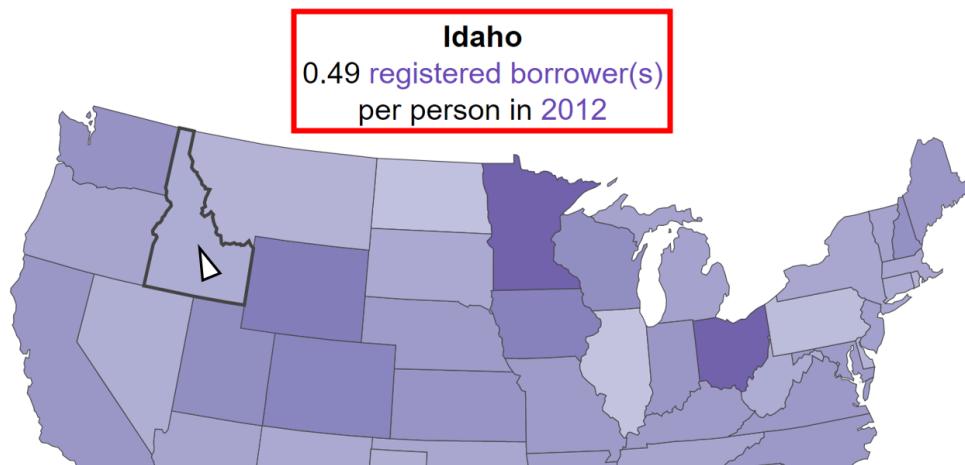


Variable: whether data is normalized by population or not

Range of values: Total count, Normalized by population

Effect: allow toggling between normalized and unnormalized data

Hover over State

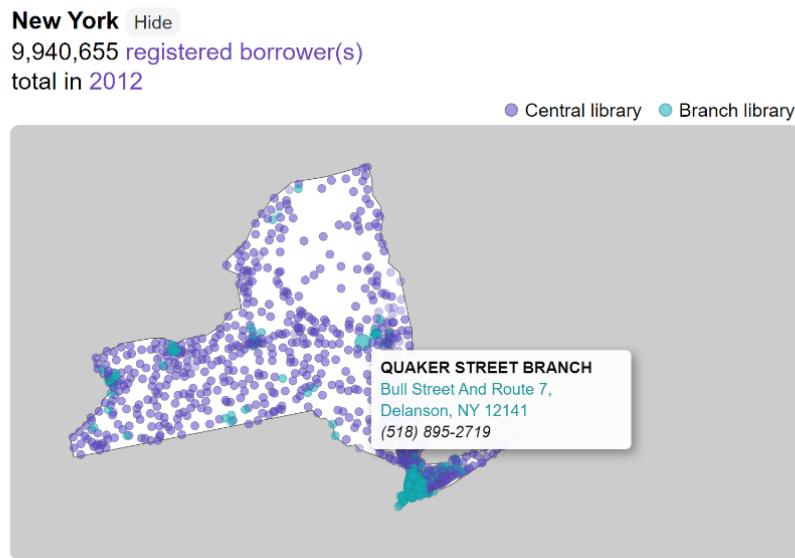


Effect: display state-specific information in tooltip at the top of the visualization

Click on State

Effect: display side panel with magnified view of state, which shows individual libraries overlaid onto the state, and state-specific line chart for field value over the years

Hover over Library



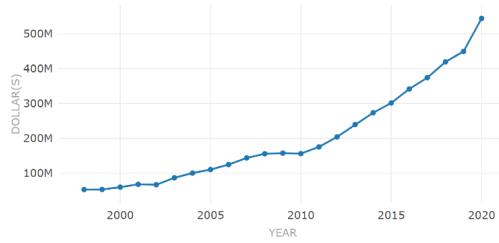
Effect: show library tooltip with library name, address, and phone number

Click on Library

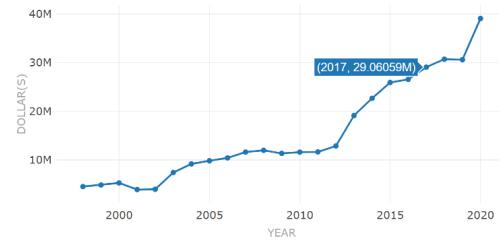
Effect: view library in Google Maps

Hover over Data Point

E-Material Expenditure in the U.S. (1998-2020)



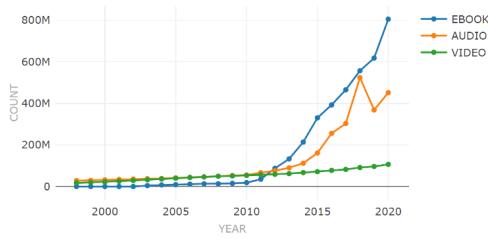
E-Material Expenditure in New York (1998-2020)



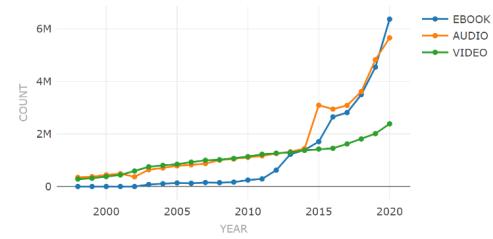
Effect: display data point coordinates (*year, value*)

Hover over Doughnut Chart

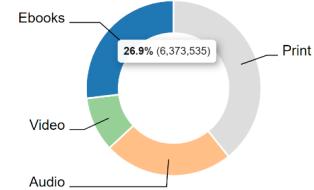
Non-Print Materials (total) in the U.S. (1998-2020)



Non-Print Materials (total) in CO (1998-2020)

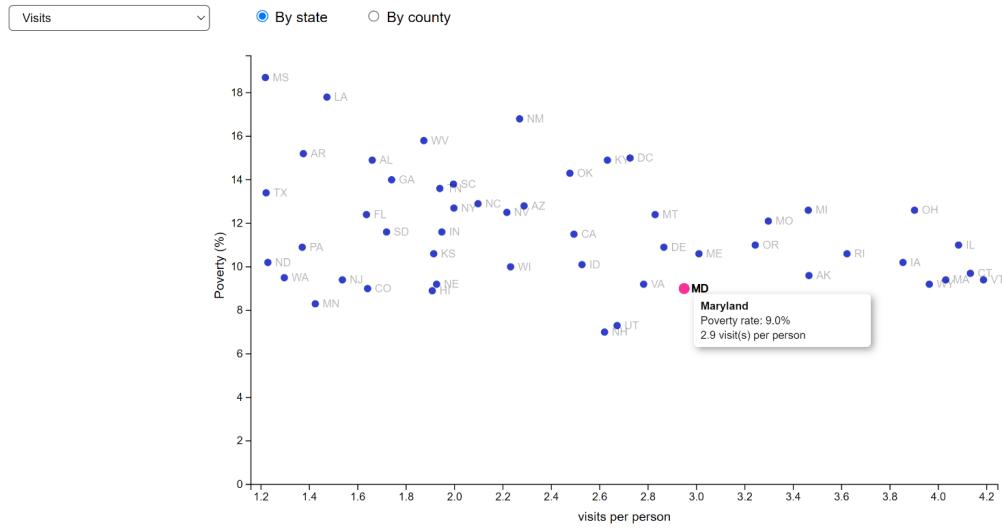


Library Materials in Colorado (2020)



Effect: display tooltip with percentage and count

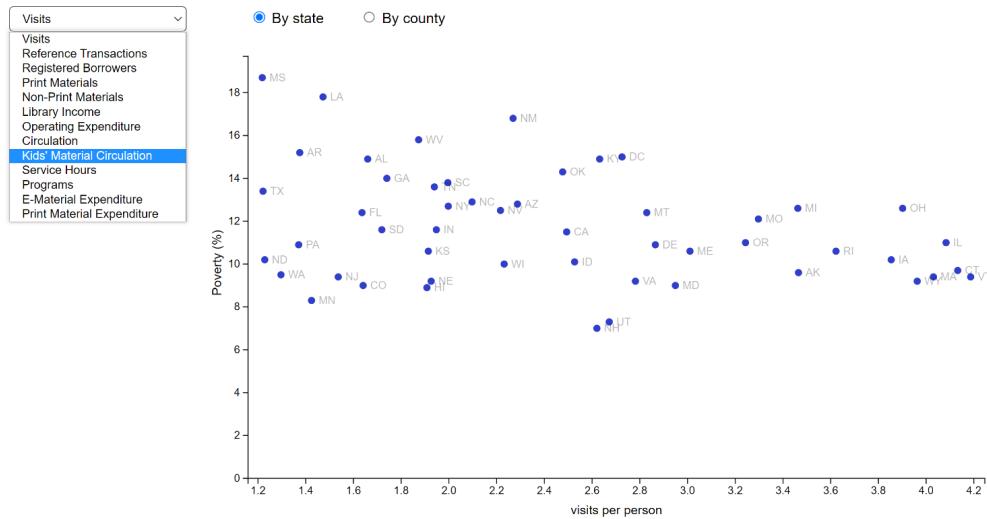
Visualization Set #2



The second set of visualizations is a scatterplot of poverty rates on the vertical (y) axis over PLS statistics on the horizontal (x) axis. This visualization set utilizes the 2020 PLS dataset and 2020 SAIPE data. There are two views: a State View (shown above) and a County View; the user can toggle between the two views by using the radio buttons above the plot. Again, there is a dropdown selection tool for users to select from various PLS statistics. Additional information is displayed on hover over data points. In State View, data points are labeled with the corresponding state abbreviation; this detail was added after observing users trying to search for a particular state of interest by exhaustively inspecting each data point.

Below I give an outline of the usage of this set of visualizations.

Dropdown



Variable: PLS data field

Range of values:

- VISITS: Visits
- REFERENC: Reference Transactions
- REGBOR: Registered Borrowers
- BKVOL: Print Materials
- OTHMAT: Non-Print Materials
- TOTINCM: Library Income
- TOTOPEXP: Operating Expenditure
- TOTCIR: Circulation
- KIDCIRCL: Kids' Material Circulation
- HRS_OPEN: Service Hours
- TOTPRO: Programs
- ELMATEXP: E-Material Expenditure (only enabled for State View)
- PRMATEXP: Print Material Expenditure (only enabled for State View)

Effect: configure visualization using selected statistic

Radio Buttons

Variable: State View versus County View

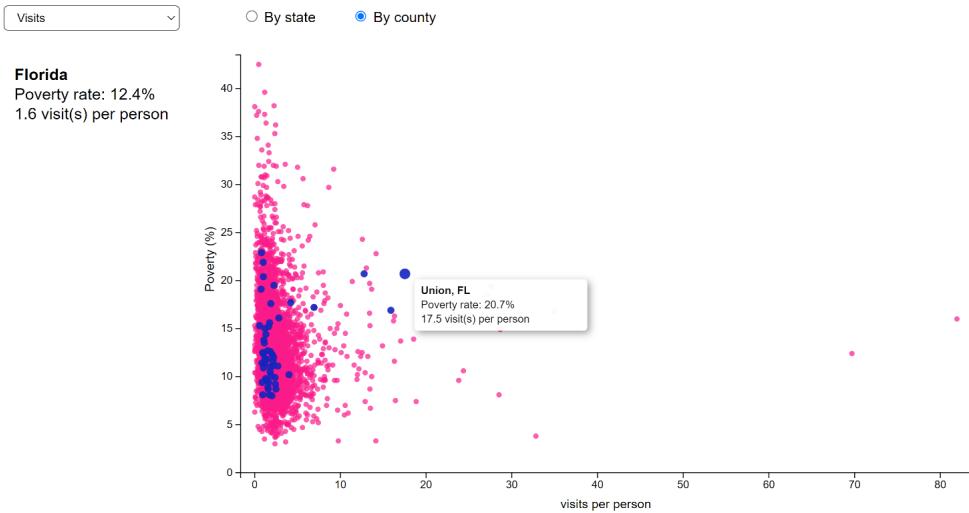
Range of values: By state, By county

Effect: allow toggling between State and County views

State View: Hover over Data Point

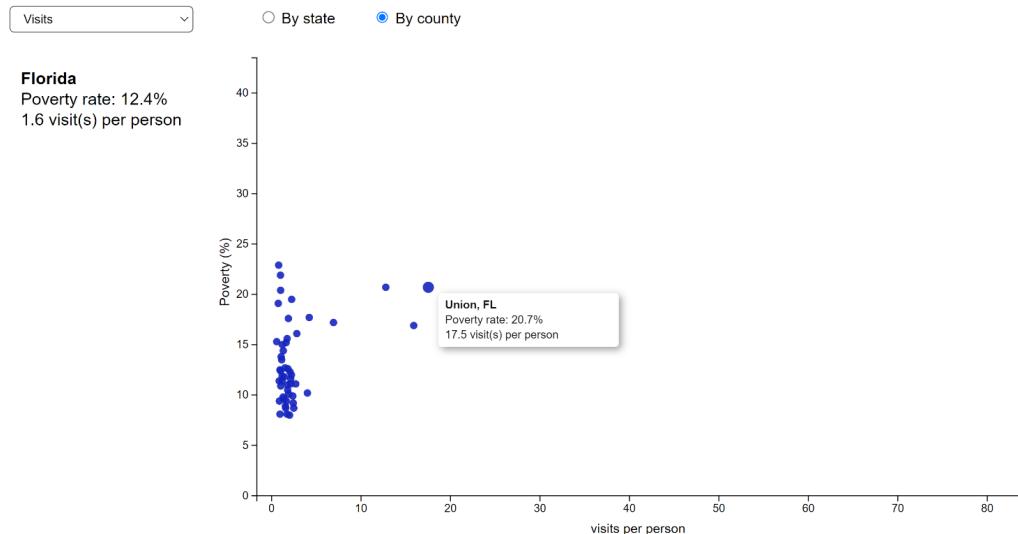
Effect: display tooltip

County View: Hover over Data Point



Effect: display tooltip, highlight all other data points for counties in the same state, and displays state-level information on left side panel

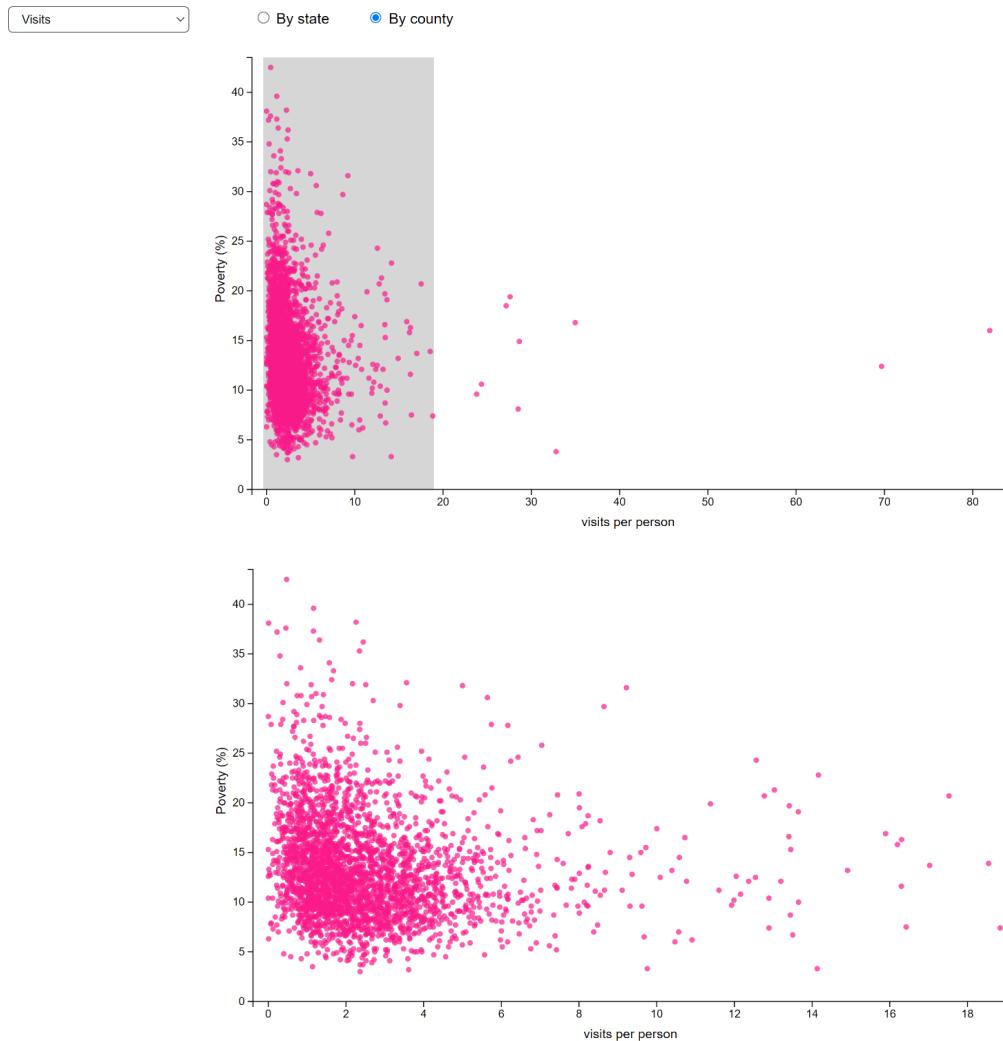
County View: Click on Data Point



Effect: toggle between two views:

- 1) hide all data points for counties not in the same state, fixes state-level information on left side panel, and
- 2) restore other data points and unfixes state-level information on left side panel

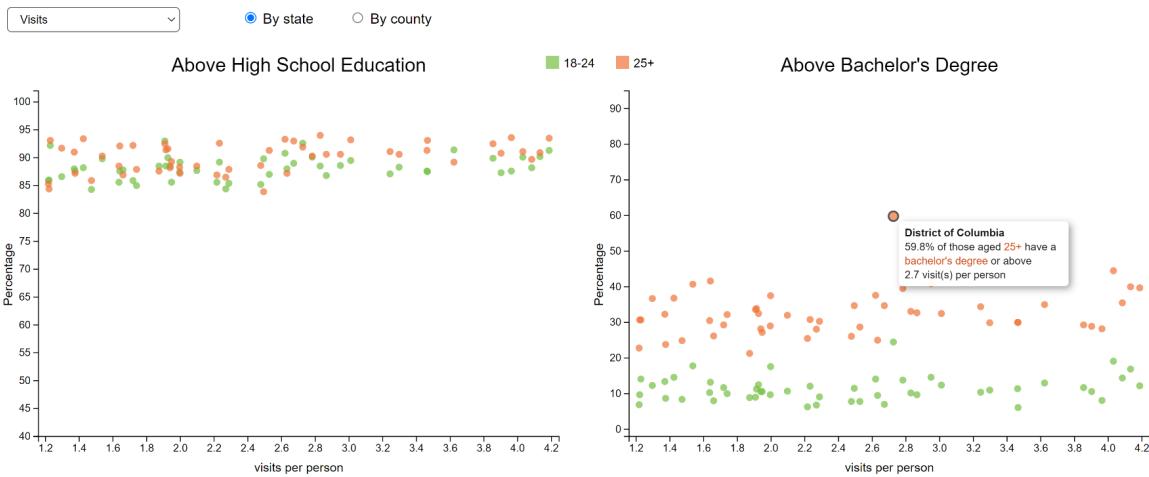
Brush Selection



Usage: hold down and drag to select an area to zoom; double click to reset to default

Effect: rescale x-axis to selected portion, e.g. to exclude any outliers

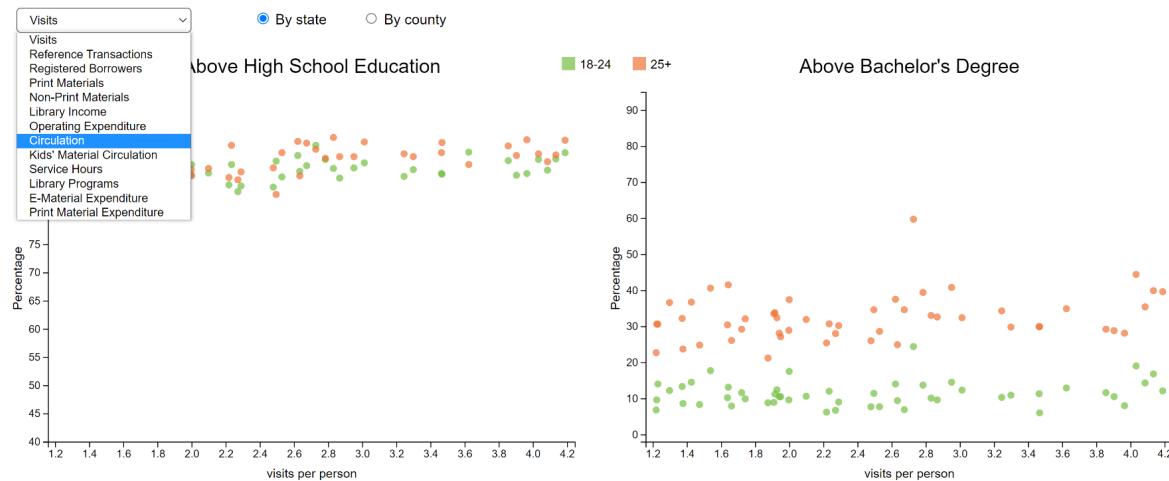
Visualization Set #3



The third set of visualizations contains two scatterplot areas side-by-side and explores the correlations between PLS statistics (x-axis) and education attainment (y-axis). This set of visualizations uses the 2020 PLS dataset and 2020 ACS data on education attainment. The left plot plots the percentage of people with a high school diploma or above (y-axis) against a particular PLS statistic (x-axis); the right plot plots the percentage of people with a bachelor's degree or above (y-axis) against a particular PLS statistic (x-axis). Like Visualization Set #2, there are two views: a State View (shown above) and a County View; the user can toggle between the two views by using the radio buttons above the plot. Again, there is a dropdown selection tool for users to select from various PLS statistics. Additional information is displayed on hover over data points. Color is used to encode age group (18-24 versus 25+). The user can choose to hide a particular age group if they so wish.

Below I outline the usage for this set of visualizations.

Dropdown



Variable: PLS data field

Range of values:

- VISITS: Visits
- REFERENC: Reference Transactions
- REGBOR: Registered Borrowers
- BKVOL: Print Materials
- OTHMAT: Non-Print Materials
- TOTINCM: Library Income
- TOTOPEXP: Operating Expenditure
- TOTCIR: Circulation
- KIDCIRCL: Kids' Material Circulation
- HRS_OPEN: Service Hours
- TOTPRO: Programs
- ELMATEXP: E-Material Expenditure (only enabled for State View)
- PRMATEXP: Print Material Expenditure (only enabled for State View)

Effect: configure visualization using selected statistic

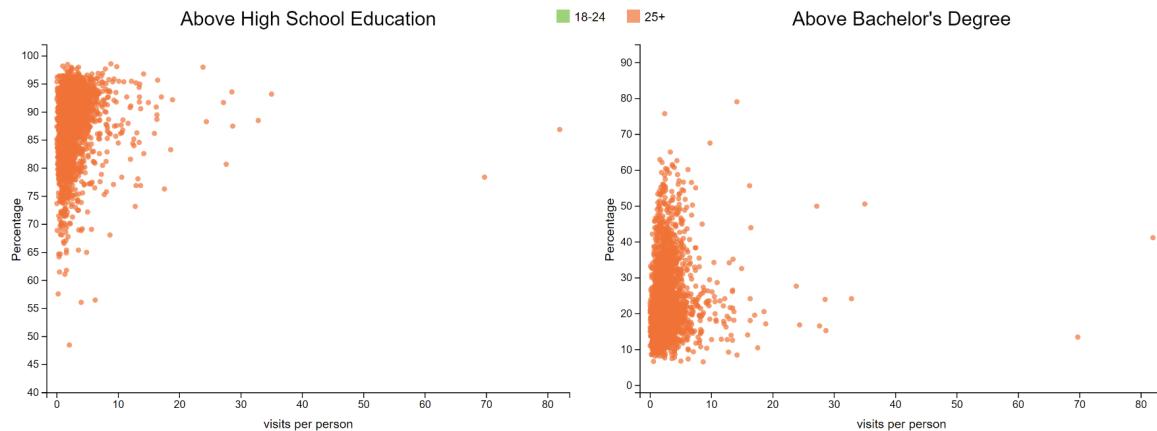
Radio Buttons

Variable: State View versus County View

Range of values: By state, By county

Effect: allow toggling between State and County views

Click on Legend

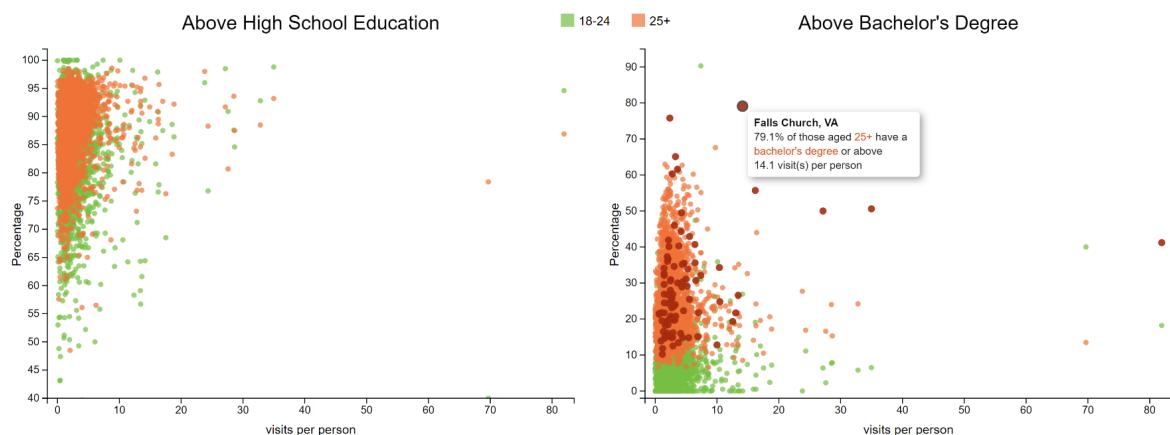


Effect: allows user to hide and show a particular age group

State View: Hover over Data Point

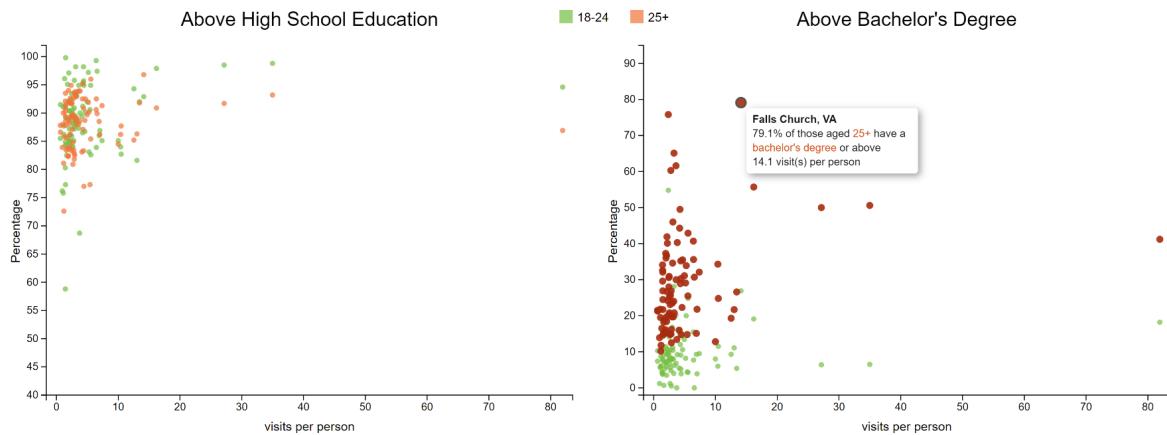
Effect: display tooltip

County View: Hover over Data Point



Effect: display tooltip, highlight all other data points in the same age group for counties in the same state

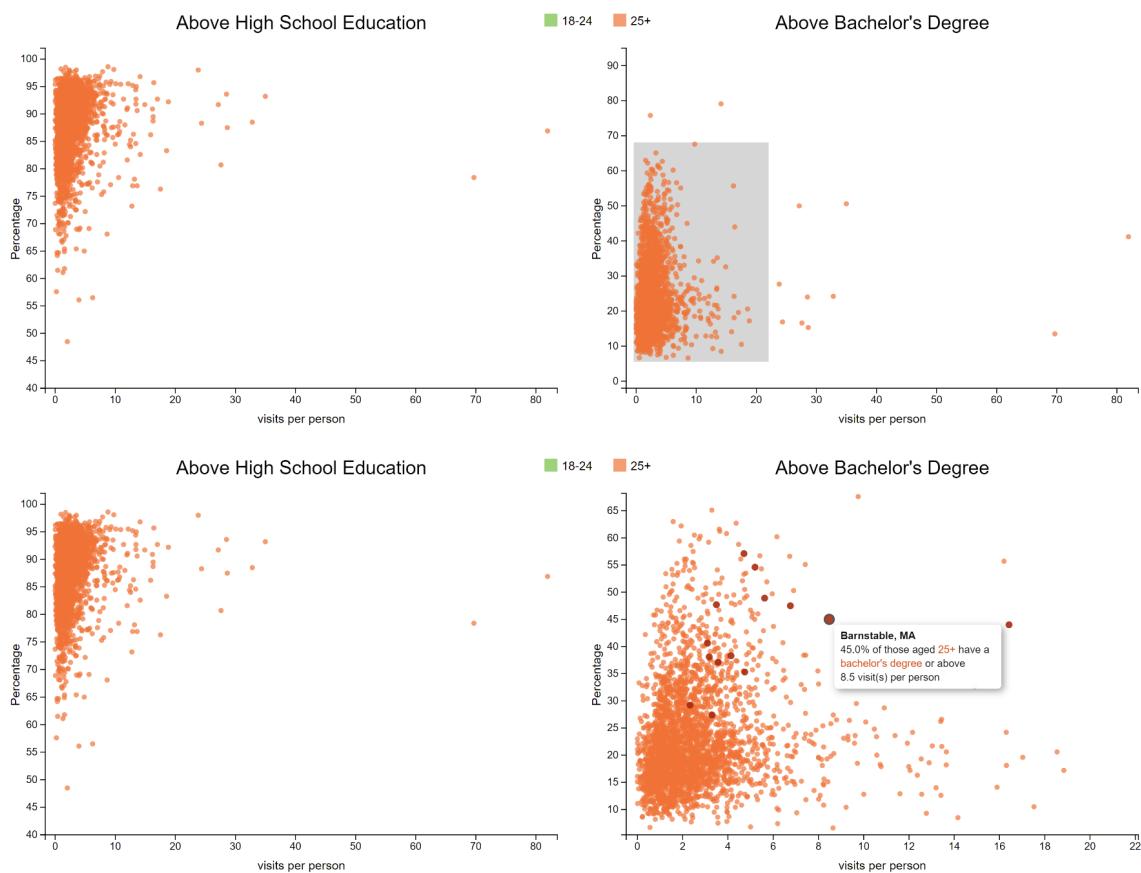
County View: Click on Data Point



Effect: toggle between two views:

- 3) hide all data points for counties not in the same state, and
- 4) restore other data points

Brush Selection



Usage: hold down and drag to select an area to zoom; double click to reset to default

Effect: rescale x- and y-axis to selected portion, e.g. to exclude any outliers or for clarity

Observations

The number of library visits in the U.S. (both per person and in total) appears to have peaked in 2009, with 1,591,292,559 visits total and 5.19 visits per person in 2009. In contrast, we can clearly see the impact of COVID-19 in the sudden drop of many statistics in 2020, such as number of visits, circulation, transactions, and library service hours. For example, there are only 732,356,964 visits total and 2.21 visits per person in 2020, less than half of the values in 2009.

Visits (total) in the U.S. (1998-2020)

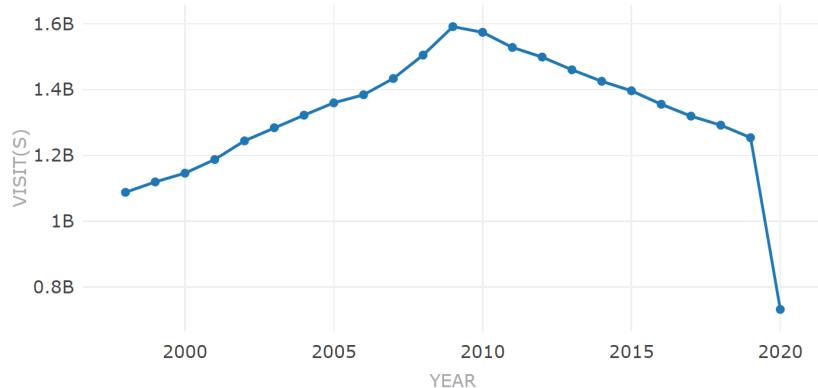


Figure 1

The number of total print materials owned by public libraries also peaked at around 2008-2009, and has thereafter shown a downward trend, as shown below in Figure 2.

Print Materials (total) in the U.S. (1998-2020)

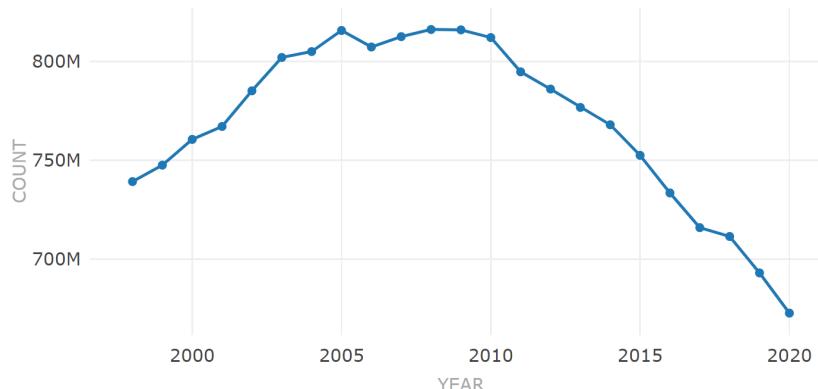


Figure 2

Unsurprisingly, the number of non-print materials (i.e. ebook, video, audio), on the other hand, has been increasing, as in Figure 3.

Non-Print Materials (total) in the U.S. (1998-2020)

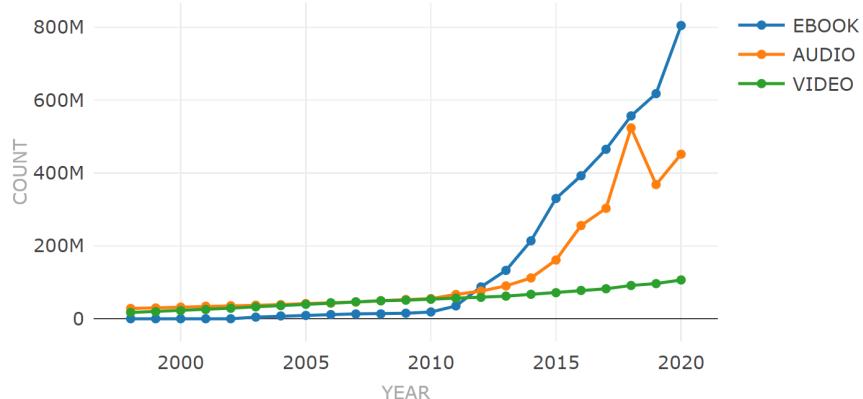


Figure 3

Interestingly, Kansas seems to have the highest number of non-print materials per person in 2020, at 54.8 items per person while the average for the entire nation is 4.11 items per person (see Figure 4).

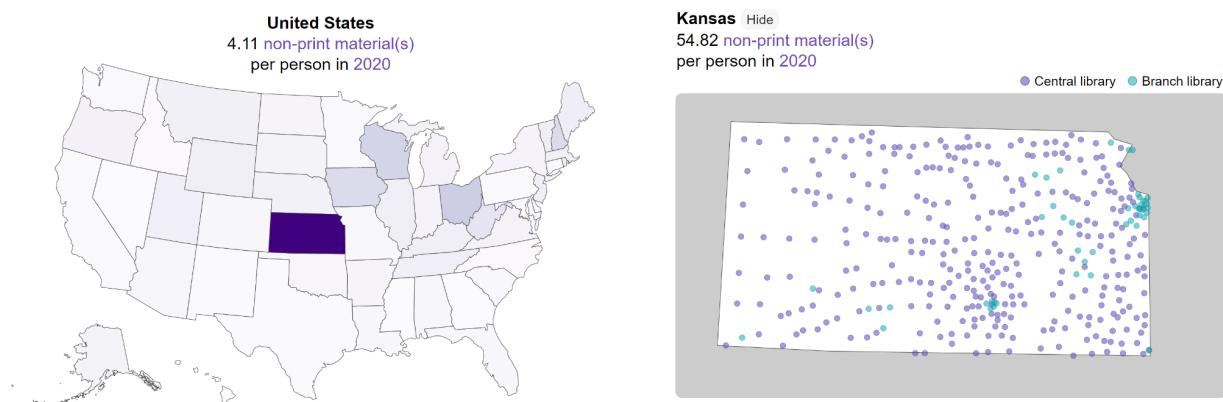


Figure 4

There are at least two possible explanations for this. First, it could be that in 2020, perhaps in response to the outbreak of COVID-19, public libraries in Kansas drastically increased their holdings of digital materials. Comparing data from previous years (e.g., 2017, 2018, 2019), we note that Kansas appears to already have had higher values for non-print materials per person than the national average (a ratio of around 6:1) in years prior to 2020. But while it is not impossible that they increased the size of their holdings in response to the pandemic to arrive at 54.82 items per person, there is also the second possibility of mistakes in data entry or data collection on the part of public libraries filling out the IMLS's Public Libraries Survey. Support for

the second explanation comes from the observation that, in 2020, Kansas did not seem to have spent significantly more on e-materials than the national average (see Figure 5).

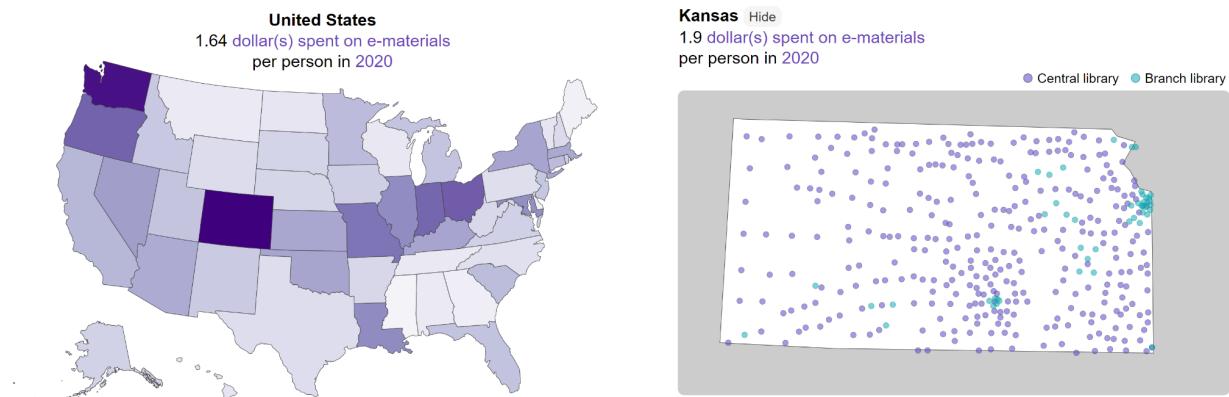


Figure 5

Once data for 2021 and later years are out, we will be able to determine with even more certainty which of the two explanations is correct.

From 1998 to 2020, the two states in which public libraries have received the greatest amount of funding appear to be California and New York, at approximately \$1.8 billion and \$1.5 billion, respectively, in 2020. Correspondingly, the total operating expenditures of public libraries in those two states are also the highest, at approximately \$1.7 billion and \$1.3 billion in 2020, respectively.

In 2020, at the level of states, we see a very weak negative correlation between poverty rate and statistics such as per capita library visits, size of holdings (e.g. print materials), library programs, and circulation (especially that of children's materials).

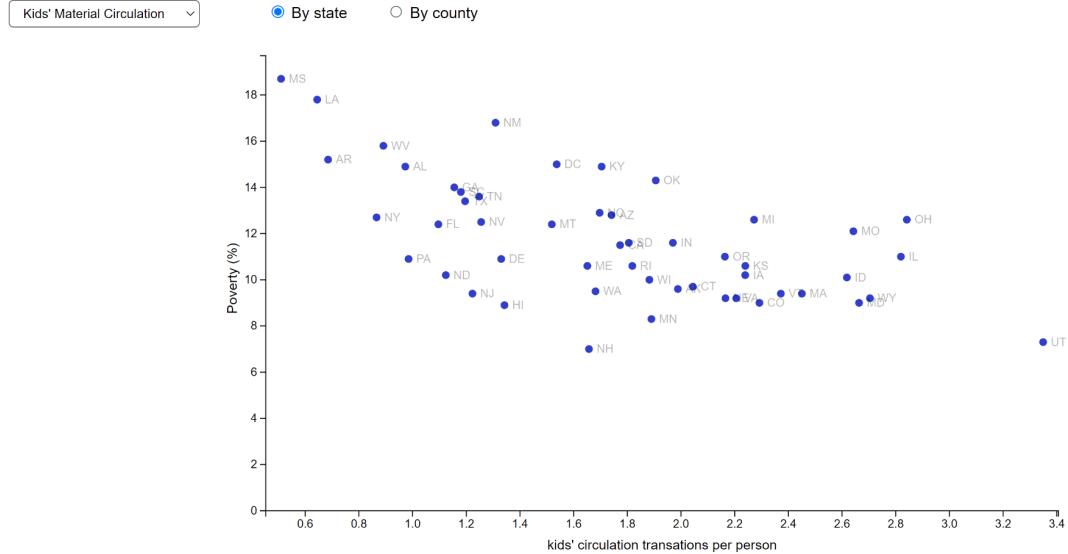


Figure 6

We also find similarly weak positive correlations between education attainment and these statistics at the level of states. When looking more specifically at counties within a particular state, however, we find more variation, depending on the particular state. For example, the number of library visits per person in counties in Texas in 2020 (Figure 7) does not seem to reveal any correlation.

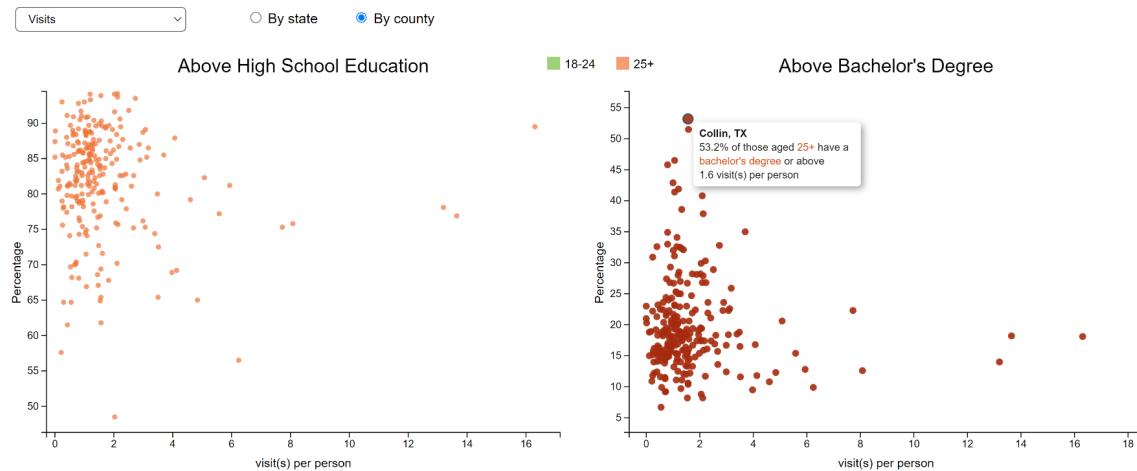


Figure 7

Future Directions

As PLS data from 2021 and onwards are made available by the IMLS, these visualizations can be updated. It would be especially interesting to see what becomes of public libraries in the

post-COVID U.S.; whether the transition from in-person to virtual activities will have had any impact on the functions of public libraries; and whether the sudden drop in the number of library visits in 2020 will ever recover, or if the pandemic will be the start of a new era, one in which in-person library visits are replaced by virtual interactions, such as the checking out of e-materials online.

Additionally, as the poverty and education visualizations (sets #2 and #3) are based on data from 2020, a year for which PLS and census data may not be entirely representative of the status of public libraries and poverty rates, it may be interesting to update these visualizations and explore these statistics for other years.

Because this project makes use of static data files, for the sake of efficiency, I made the conscious decision to limit the size of these data files such that they will be able to load in a reasonable amount of time. In the future, it may be helpful to store these data in a database and query the necessary data from the database as needed so that visualizations, in particular those that involve county-level data, can be configured to allow users to explore even more aspects of the datasets.

Acknowledgements

Thank you to Professor Holly Rushmeier, for introducing me to the world of data visualization in CPSC 446, and for supporting this idea and me throughout the course of this project, and for loving public libraries too.

References

The Complicated Role of the Modern Public Library

Jennifer Howard. *Humanities*.

<https://www.neh.gov/article/complicated-role-modern-public-library>

To Restore Civil Society, Start With the Library

Eric Klinenberg. 2018. *The New York Times*.

<https://www.nytimes.com/2018/09/08/opinion/sunday/civil-society-library.html>

Most Americans – especially Millennials – say libraries can help them find reliable, trustworthy information

A. W. Geiger. 2017. *Pew Research Center.*

<https://www.pewresearch.org/fact-tank/2017/08/30/most-americans-especially-millennials-say-libraries-can-help-them-find-reliable-trustworthy-information/>

Public Libraries Survey

Institute of Museum and Library Services.

<https://www.imls.gov/research-evaluation/data-collection/public-libraries-survey>

Small Area Income and Poverty Estimates (SAIPE) Program

Census.gov. <https://www.census.gov/programs-surveys/saipe.html>

American Community Survey

Census.gov. <https://www.census.gov/programs-surveys/acs>

Interactive Data Visualization for the Web

Scott Murray. 2017. 978-1-491-92128-9

Slider Range Mars Weight Calculator

Marine Piette. *CodePen.* <https://codepen.io/mayuMPH/pen/ZjxGEY>

Donut chart with group label in d3.js

Yan Holtz. 2018. *The D3.js Graph Gallery.*

https://d3-graph-gallery.com/graph/donut_label.html

Zooming in d3.js

Yan Holtz. 2018. *The D3.js Graph Gallery.*

https://d3-graph-gallery.com/graph/interactivity_zoom.html