Project Progress Report 2

(due May 24th 11:59p.m)

You can start working on the project once your report is accepted and graded by your TA. The entire final project is worth **35%** of your final grade and this report accounts for **10%**. This project is done individually.

Submission Guideline

Download this google doc, fill the table. **Type** your answers, no handwritten answers will be accepted (except for the very last question). Submit it in **PDF** format on Gradescope.

If you need some inspirations please feel free to take a look at:
Showcase of Information is Beautiful Awards
Bloomberg Year In Graphics Review
The Pudding
The New York Times

Project Guidelines

Note: The guideline has been further clarified from Progress Report 1, so double-check whether your dataset choice still satisfies the updated guideline below.

1. You may use more than one dataset, however, regardless if you use one or multiple datasets, your visualizations must make use of at least three following data types - **link, position, and attribute.**
2. You cannot use any dataset from the class (Labs, Assignments, Lecture Exercises)
3. You can make your own dataset (Web scrape etc.) provided point 1. is satisfied.

Part 1 - Story and Narrative

Link to the dataset	https://www.kaggle.com/datasets/imls/public-libraries?select=libraries.csv
Example item from the dataset	NY NY0001 MONTOUR FALLS MEMORIAL LIBRARY MONTOUR FALLS SCHUYLER 18479.0 1711.0 1698.0 1.31 0.15 1.46 11177.0 9549.0 171.0 2439.0 1218.0 17.0 11892.0 1013.0, [PQ, Santee, El Cajon], [PQ, Santee]
Story you want to deliver	(a story should be in a form of a list of facts, insights, and messages - refer to the lecture slide) I plan to focus on libraries that are within the state of New York as a lot of the data is from this state and provides for a more focused area. I will focus only on the top three counties in New York that have the most number of libraries. The story that I am trying to deliver with this dataset is that the population of a county within the state of New York reflects the number of resources available for people to use through the library/the income of the library itself. A node link diagram would allow for the visualization of all the different libraries within a county that offer a specific collection. It is expected that counties with smaller populations would have fewer links between libraries that offer the same collections as there are fewer users. In summary, the story is that as the population size of a county grows, it's likely that not only will there be more libraries within the region, but also there will be more resources available to these libraries. New York is the state with the most libraries within it, and the population in which these libraries serve is largely based on service population and the county population. Such information can indicate how many resources are given towards libraries. Also, looking at the general county resources and comparing the different libraries within the county can be quite insightful.
Describe your target audience.	(using the questions the lecture slide listed) The intended audience would be anyone interested in looking at how county size affects the amount of resources libraries receive in general. Also, another set of audience would be

	anyone who is interested in looking at the different resources that New York libraries have to offer. If the audience is not familiar with libraries as well as the availability of libraries within New York's top three counties, the introduction will introduce the topic and the interactive parts of the graph will provide more information. The audience should care as they are interested in exploring any trends that may be noticeable between the county libraries. I want the audience to take away a better understanding of the libraries within the counties that have to offer resource wise, and how the libraries compare with each other as well as the other libraries in other countries. I am assuming that they know nothing about this project besides the basics, so I will include any necessary information in the introduction and paragraph explanation parts. There will be a lot of interactivity opportunities throughout this project as they can explore each visualization and learn more about the counties and their libraries. There is not too much math going on in this project and if they are interested in looking at certain numbers, the interactivity in the graphs all for this to happen. This project would be work best on a computer for better interactivity with the graphs.
The goal of your project outcome. And why?	(exploratory vs. explanatory) The goal of this project would be for exploratory purposes. It would explore the different resources that are available to the top three counties within the state of New York, based on their population size. Such exploration would be done by allowing for the users to choose different counties they would want to explore and provide them with data visualization that would include information about the county as well as the libraries within the county. Such information would help guide their exploration and discover new trends that are present among different counties.
Narrative structure you plan to use	Martini-glass Structure

Elaborate your choice of narrative structure.	The introduction would be providing the purpose of the project, why this specific dataset, and any contextual information that would be needed to better the exploration process. Basic information about the different attributes that could be explored is necessary to present; this part can include basic plot about general information about libraries in New York and provide general statistics. Once this information is presented, the user would be able to explore information about different counties in New York through various visualizations.
Narrative genre you plan to use	Magazine Style
Elaborate your choice of narrative genre.	This project would be a blog post type of genre, including information throughout the process, but also wrapping it with text so it's clear what is going on. This would make the most sense in this scenario as the information that would be provided as the context would help lead the project along. As the user goes through the website, they will have guiding information about the graphics that are presented, but will also have a chance to explore the graphics that are presented as they go through the project.

Part 2 - Outline

Story you want to deliver	(you can copy/paste from Part 1) I plan to focus on the top three counties with (decided to narrow it down because there were too many counties within New York, but many did not have enough libraries data points to visualize enough information) libraries that are within the state of New York as a lot of the data is from this state and provides for a more focused area. The story that I am trying to deliver with this dataset is that the population of a county within the state of New York reflects the number of resources available for people to use through the library/the income of the library itself. A node link diagram would allow for the visualization of all the different libraries within a county that offer a ~~specific~~ print collection. It is expected that counties with smaller populations would have fewer links between libraries that offer the same collections as there are fewer users. In summary, the story is

	that as the population size of a county grows, it's likely that not only will there be more libraries within the region, but also there will be more resources available to these libraries. New York is the state with the most libraries within it, and the population in which these libraries serve is largely based on service population and the county population. Such information can indicate how many resources are given towards libraries. Also, looking at the general county resources and comparing the different libraries within the county can be quite insightful.
Specifications on each plot **in the order of how you lay out on your project**	(for each plot, include 1) clear task abstraction, 2) attributes used, 3) marks, 4) channels, and 5) how this plot adds to the story) ***INFORMATION IS IN TABLE BELOW***
Elaborate the choice of their marks and channels for each vis	***INFORMATION IS IN TABLE BELOW***

1. Plot 1 - ~~Network Plot on Map~~ Plotting Locations on Map
    1) Task: ~~This plot looks at the links between libraries within a library that share the same collection.~~ The plot looks at the geographic location of all the libraries for each country on the map of New York
    2) Attributes: latitude of each library, longitude of each library
    3) Marks: Area, point, ~~lines~~
    4) Channels:
        - Position
            - Point placed on longitude and latitude on a map
        - Color
            - Corresponds to which county a point belongs to
        - ~~Size~~

- ~~The service size for each library~~

5) How this plot adds to the story:
This visualization looks at a geographical map of New York, and for each county, the different locations of all the libraries within the three counties that are color coded based on the county the library belongs to.~~different links that are present for each collection~~. This would further allow exploration of understanding the distribution the libraries within~~the different offerings of resources of different counties in~~ New York, providing visual placement of how close the libraries are within a county. Such information can allow one to visually see how many libraries are within the county.~~ Also, the size of the node can indicate the likeliness of a library having a collection, as a library serving larger service populations are more likely to have a collection.~~

~~1. Plot 1 - Bar Chart (x axis → collection type, y axis → number of collections)~~

~~6) Task: This chart analyzes the overall pattern of collections that are available within libraries at the county. With this information, I can identify the more popular collection types.~~

~~7) Attributes: Collection Type, Collection Type Count~~

~~8) Marks: line mark~~

~~9) Channels:~~

- ~~Position~~
  - ~~aligned horizontal channel to match collection (collection is a category, which can only be on the horizontal axis in a bar chart)~~
  - ~~Aligned vertical channel to match total count~~

~~10) How this plot adds to the story:~~
~~This visualization would provide general overall information about the collections that are offered within these counties and the number of libraries that offer this collection. This would provide insight as to what collections to expect to find in a county.~~

I took this chart from my final project as I realized it provided repetitive information and didn't add that much value to the exploration project. The stacked bar chart that I included instead provides more information and there is more interactivity involved within it.

2. Plot 2 - Stacked Bar Chart (x axis → Library, y axis → Collection Count, Key → Collections)

11) Task: This chart analyzes trends between libraries and the number of each collection it has. This chart also allows for a comparison to be made amongst the different libraries within the county.

12) Attributes: Collection Type, Collection Type Count, Library Name

13) Marks: line mark

14) Channels:
    - Position
        - aligned horizontal channel to match Library (library is a category, which can only be on the horizontal axis in a bar chart)
        - Aligned vertical channel to match count of collection
    - Color
        - Color of the collection type
15) How this plot adds to the story:
    This visualization builds off of the previous chart, allowing for a better visualization of the accessible collections at specific collection types, and can allow for comparisons to be made.

~~3. Plot 3 - Histogram Chart (x axis → Total Number of Employee Hours, y axis → Frequency)~~

~~16) Task: This chart analyzes the number of libraries within a county that has a certain number of total hours employees are employed.~~

~~17) Attributes: Employee hours, count~~

~~18) Marks: line mark~~

~~19) Channels:~~

~~- Position~~

~~- aligned horizontal channel to match bin of number of hours~~

~~- Aligned vertical channel to match count of number of libraries (frequency can only be as a vertical position in a histogram)~~

~~- Color~~

~~- Color of the collection type~~

~~20) How this plot adds to the story:~~

~~This visualization would support general trends that would be present amongst employed hours and county population. It is expected that counties that have lower population would have few employees working full time and less total number of employee hours in general. This would allow further exploration about how county size affects resources.~~

~~4. Plot 4 - Histogram Chart (x axis → Total Number of Librarians, y axis → Frequency)~~

~~21) Task: This chart analyzes the number of libraries within a county that has a certain of librarians working (per 25000 in service population)~~

~~22) Attributes: Librarians, count~~

~~23) Marks: line mark~~

~~24) Channels:~~

~~- Position~~
~~- aligned horizontal channel to match bin of number of librarians~~
~~- Aligned vertical channel to match count of number of libraries (frequency can only be as a vertical position in a histogram)~~

~~25) How this plot adds to the story:~~

~~This visualization is similar to the previous one, but focuses only on librarians. This is important to know for the story as the librarians are essential for the running of the libraries. Similar to the previous chart, this chart would support general trends that would be present amongst the number of librarians and county population. It is expected that counties that have lower population would have fewer librarians working. This would allow further exploration about how county size affects resources.~~

I took these two chart outs from my final project because I realized the information can be provided to the users in a simpler way through interactivity on another plot. This allows for more direct information being portrayed regarding a library and more exploration to be involved.

3. Plot 3 - Scatter Plot (x axis → Total Number of Librarians, y axis → Frequency)

26) Task: This chart analyzes the trend between the number of visits from libraries in a county and the number of registered users.

27) Attributes: Number of Library Visits, Number of Registered Users

28) Marks: point mark

29) Channels:
- Position
  - aligned vertical channel to the number of registered user
  - Aligned horizontal channel to the number of library visits
- Size
  - The service size for each library

30) How this plot adds to the story:

This visualization provides information about trends between library visits and registered users, taking into account service population as well. Users can explore which libraries could seem like outliers based on their position on the graph as well as look at how service population affects registered users and library visits.

4. Plot 4 - Network Diagram

31) Task: This plot looks at the links between libraries within a library that share the print collection.

32) Attributes: Link with other library based on collection~~, latitude, longitude~~

33) Marks: ~~Area,~~ point, lines

34) Channels:

a) Color → when hovered over, then will show direct link between two nodes

~~- Position~~

~~- Point placed on longitude and latitude on a map~~

~~- Size~~

~~- The service size for each library~~

35) How this plot adds to the story:

a) This visualization looks at a network diagram that connects all the libraries within the same county that also have a print collection. This way, users who are exploring this project are able to locate libraries that are connected and can look for other print collections at the different libraries if one of them doesn't have a collection. This visualization also provides information about how all the libraries within the county can be connected in some way and can share the same resources.

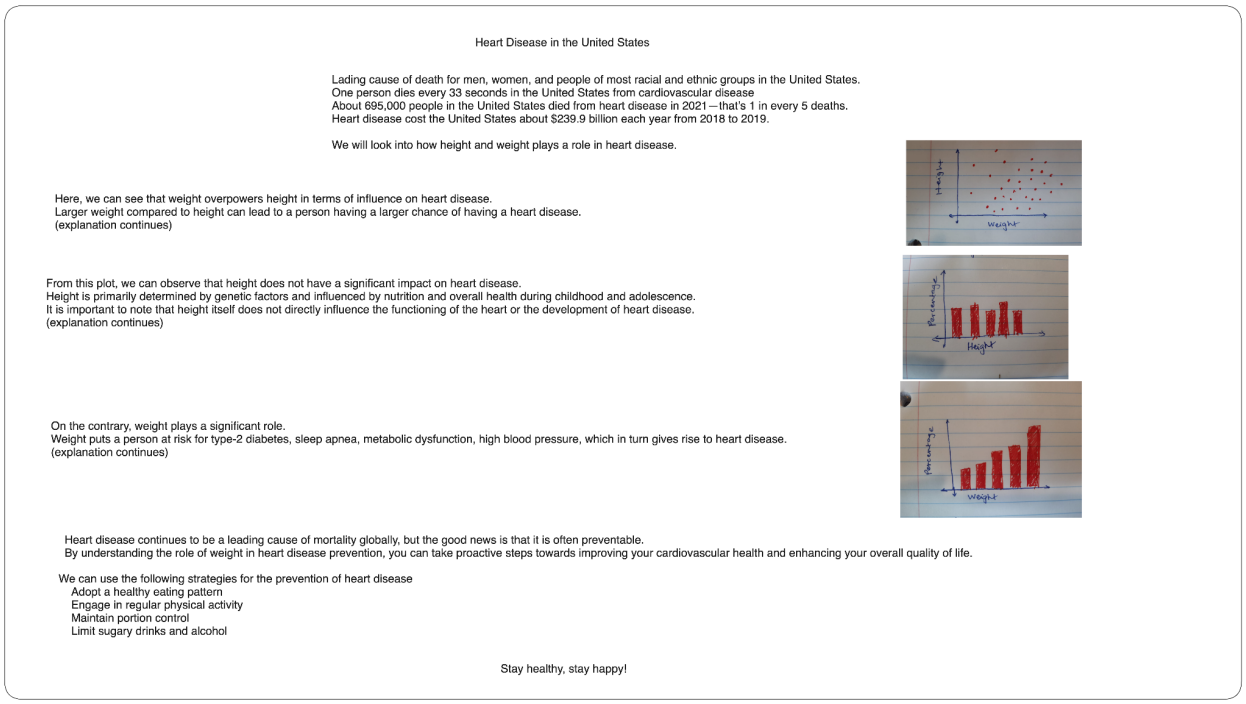
~~This visualization looks at a geographical map of New York, and for each county, the different links that are present for each collection. This would further allow exploration of the different offerings of resources of different counties in New York, providing visual placement of how close the libraries are within a county. Also, the size of the node can indicate the likeliness of a library having a collection, as a library serving larger service populations are more likely to have a collection.~~

Note: I did all my data cleaning in a jupyter notebook using python. I also create the relevant json files for my specific data during my data cleaning process in jupyter notebook .

Part 3 - Prototype

Provide a photo or screenshot of your prototype. A prototype should depict how you place different components of your visualization. You may use pen-paper, or using tools like excalidraw, figma etc.

A basic, barebones sample prototype for this project



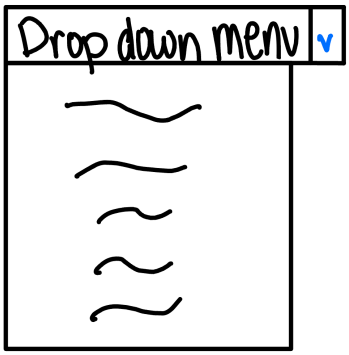
Accessible Library Resources in the State of New York

New York State has the highest number of libraries accessible within the state. The resources available at the library are greatly affected by county exploration and service population.

To give you general information about the number of libraries that in New York that caters to the people, you can find some statistics below:

- Total Number of Counties: INSERT NUMBER
- Total Population Size of New York: INSERT NUMBER
- Largest Service Population that a Library Caters to: INSERT MAX SERVICE POPULATION, NAME
- Smallest Service Population that a Library Caters to: INSERT MIN SERVICE POPULATION, NAME
-

To provide opportunities for a deeper exploration, select which county you would like to explore :

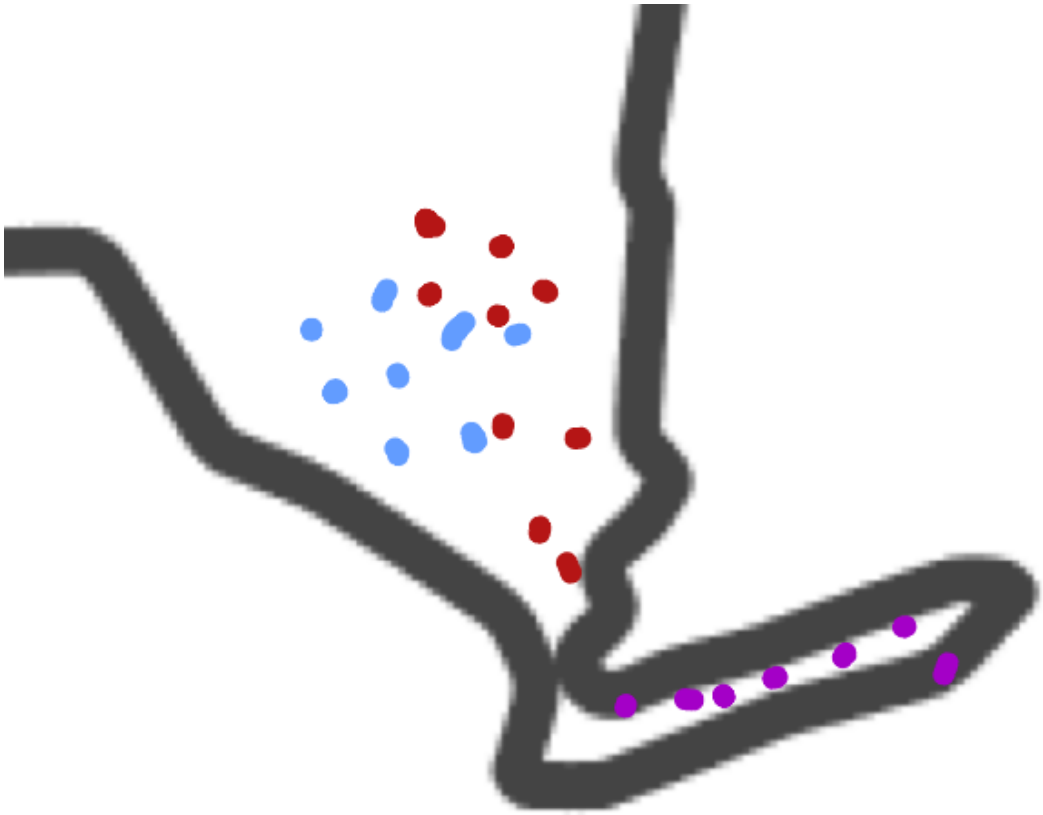


Great! Now that you've selected a county, here are some general statistics about this county.

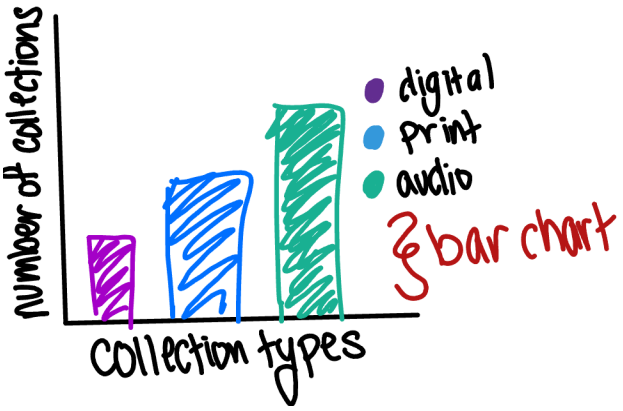
- County Population: INSERT NUMBER
- ~~Total Number of Libraries: INSERT COUNT~~
- ~~Library with the largest service population: INSERT NAME, NUMBER~~
- ~~Library with the smallest service population: INSERT NAME, NUMBER~~

(This information is now provided through interactivity)

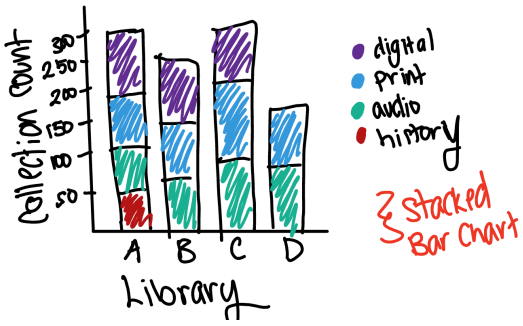
Let's take a look at how the libraries are distributed on the map within these three counties.You can over over the points to learn more information about each library.



~~To start exploring, let's look at this county's availability with collection types, counting the number of libraries within the libraries that have this county.~~ (REMOVED THIS CHART - EXPLAINED ABOVE)



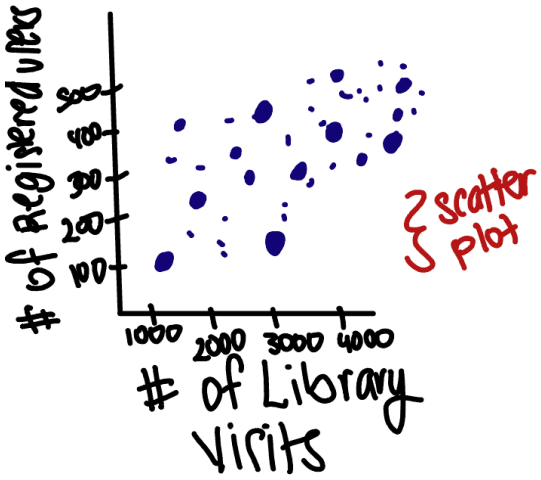
Now let's take a deep look into which libraries have which collections, and how many of each collection.



Now let's take a look at the general trend available for employees within the counties. This first chart provides information about the number of libraries within the county that meet a certain number of employee hours. ~~As you explore different counties, you will likely notice that as the county population increases, it is likely that the graph will skewed towards the right as more employees would need to be employed.~~ (REMOVED THIS CHART - EXPLAINED ABOVE)

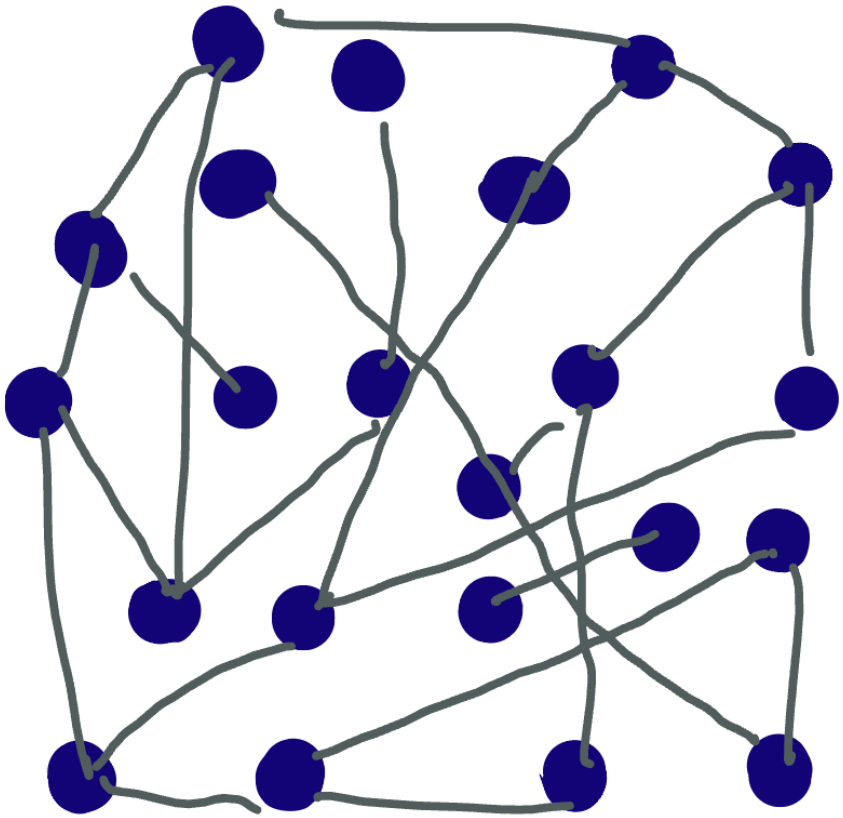
~~Similarly, let's explore librarians, as they are an essential resource to the library.~~ (REMOVED THIS CHART - EXPLAINED ABOVE)

Now let's look at the plot below, which shows a positive linear relationship trend between number of library visits and number of registered users. To further your exploration, you can go over a point and look at the service population a library caters to .



Finally, let's look how easy it is to find print collections between libraries in this county.
~~Select the collection you would like to explore. [INSERT DROP DOWN MENU]~~

This network map below shows the connection between libraries within the county that have this collection. You can hover over a certain node and see the library name. You can also hover over any link on the diagram and see which link is connected to a library.



You've completed your exploration of libraries with different counties in the state of New York! You've had a chance to look at different trends that were noticeable and possibly draw conclusions as to why these trends were present, whether because of county population or service population. Since this dataset is from 2014, the next step that would need to be done is to look at how these trends change over time and plot out the present trends of libraries in New York during the year of 2023.