

The background image shows a grand, traditional library. The ceiling is high and features a series of large, dark wood arches. The walls are filled with floor-to-ceiling bookshelves, all packed with books. In the foreground, there are rows of wooden study carrels with small, white, bust-shaped figures standing on the ledges. A green plaque on the right side of the image reads "ANNO MDCC".

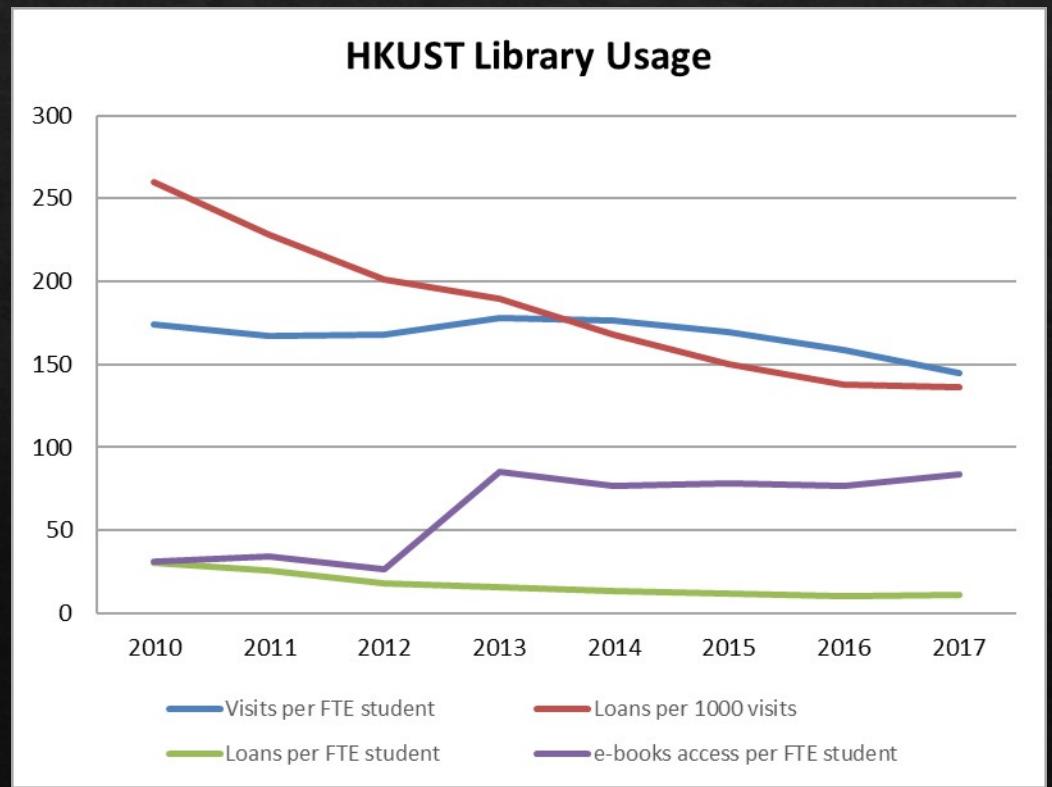
Term Project Team Four

Kris Green, Eduardo Jimenez, Shilpa Kadambi Sudarsanam, Jo Vivian



Business Problem

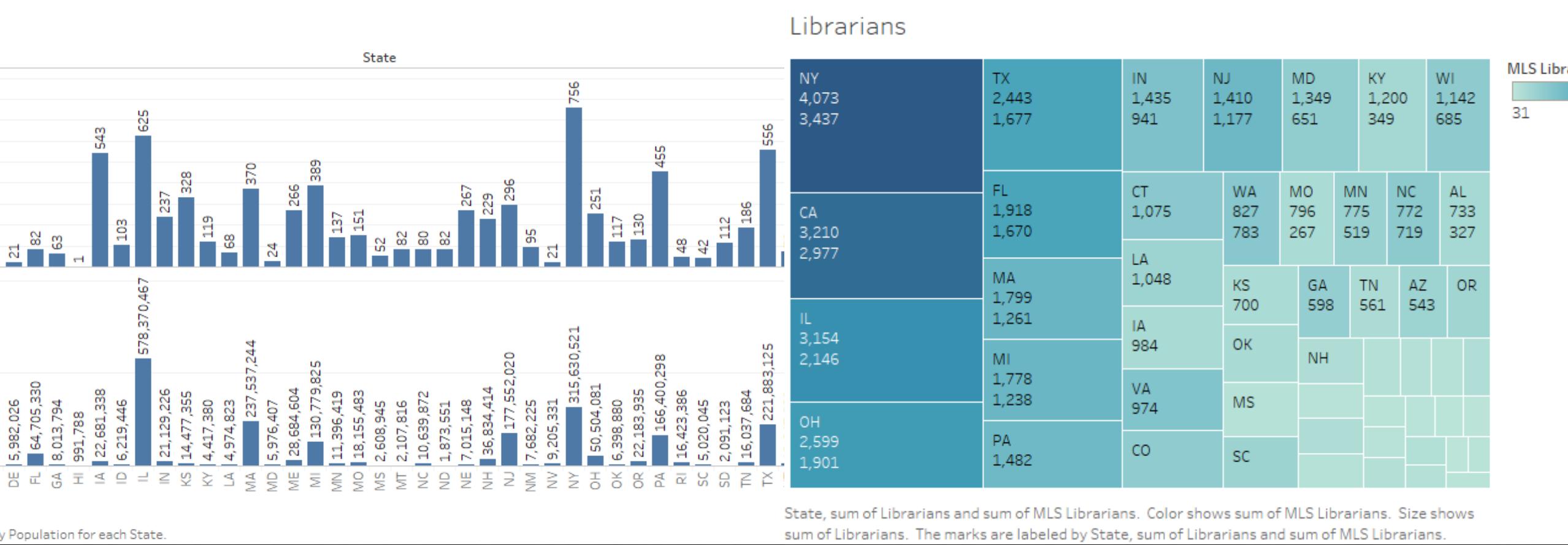
- ❖ Increase number of visits to public libraries



Data

- ❖ 2016 Public Library survey conducted by the Institute of Museum and Library Services
- ❖ Includes all public libraries identified by state library administrative agencies in the 50 states and the District of Columbia
- ❖ 75 columns x 8,813 rows

Library ID	Submission Year	Library Name	Address	City	Zip Code	Longitude	Latitude	State Code	Count
01	2015	ANCHOR F	72551 MIL	ANCHOR F	99556	-151.84	59.7796	2	
02	2015	ANCHORA	3600 DEN	ANCHORA	99503	-149.877	61.18768	2	
03	2015	ANDERSO	101 FIRST	ANDERSO	99744	-149.179	64.3436	2	
06	2015	KUSKOKW	420 CHIEF	BETHEL	99559	-161.803	60.78576	2	
07	2015	BIG LAKE F	3140 SOUT	BIG LAKE	99652	-149.818	61.54661	2	
08	2015	CANTWEL	1 SCHOOL	CANTWEL	99729	-148.757	63.46105	2	
11	2015	CHINIAK F	43318 SPR	CHINIAK	99615	-152.233	57.61223	2	
14	2015	COLD BAY	10 BARAN	COLD BAY	99571	-162.713	55.20663	2	
15	2015	COOPER L	MILE .8 BE	COOPER L	99572	-150.011	60.4874	2	
16	2015	CORDOVA	622 FIRST	CORDOVA	99574	-145.766	60.548	2	
17	2015	CRAIG PUI	504 THIRD	CRAIG	99921	-133.151	55.47613	2	
18	2015	IPNATCHI	59 MAIN S	DEERING	99736	0	0	2	
19	2015	DELTA COI	2291 DEBC	DELTA JUN	99737	-145.733	64.04657	2	
20	2015	DILLINGHA	306 D STRE	DILLINGHA	99576	-158.467	59.04025	2	
21	2015	EAGLE PUI	4TH & AM	EAGLE	99738	-141.202	64.78662	2	
23	2015	FAIRBANK	1215 COW	FAIRBANK	99701	-147.739	64.83832	2	
25	2015	CHARLES E	299 ANTO	GALENA	99741	-156.872	64.74802	2	
27	2015	COPPER V	MILE 186 C	GLENNALL	99588	-146.05	62.19841	2	
28	2015	GUSTAVU	14 GUSTAVU	GUSTAVU	99826	-135.727	58.41308	2	
29	2015	HAINES BC	111 THIRD	HAINES	99827	-135.448	59.23284	2	
30	2015	TRI-VALLE	400 SUNTF	HEALY	99743	-149.017	63.87349	2	
31	2015	HOLLIS PU	NO. 1 HOL	HOLLIS	99950	-133.076	55.78188	2	
32	2015	HOMER PU	500 HAZEL	HOMER	99603	-151.544	59.6405	2	
33	2015	HOPE CON	18487 A A'	HOPE	99605	-149.445	60.10942	2	
35	2015	JUNEAULD	202 MARIN	JUNEAUL	99801	-124.404	58.30848	2	

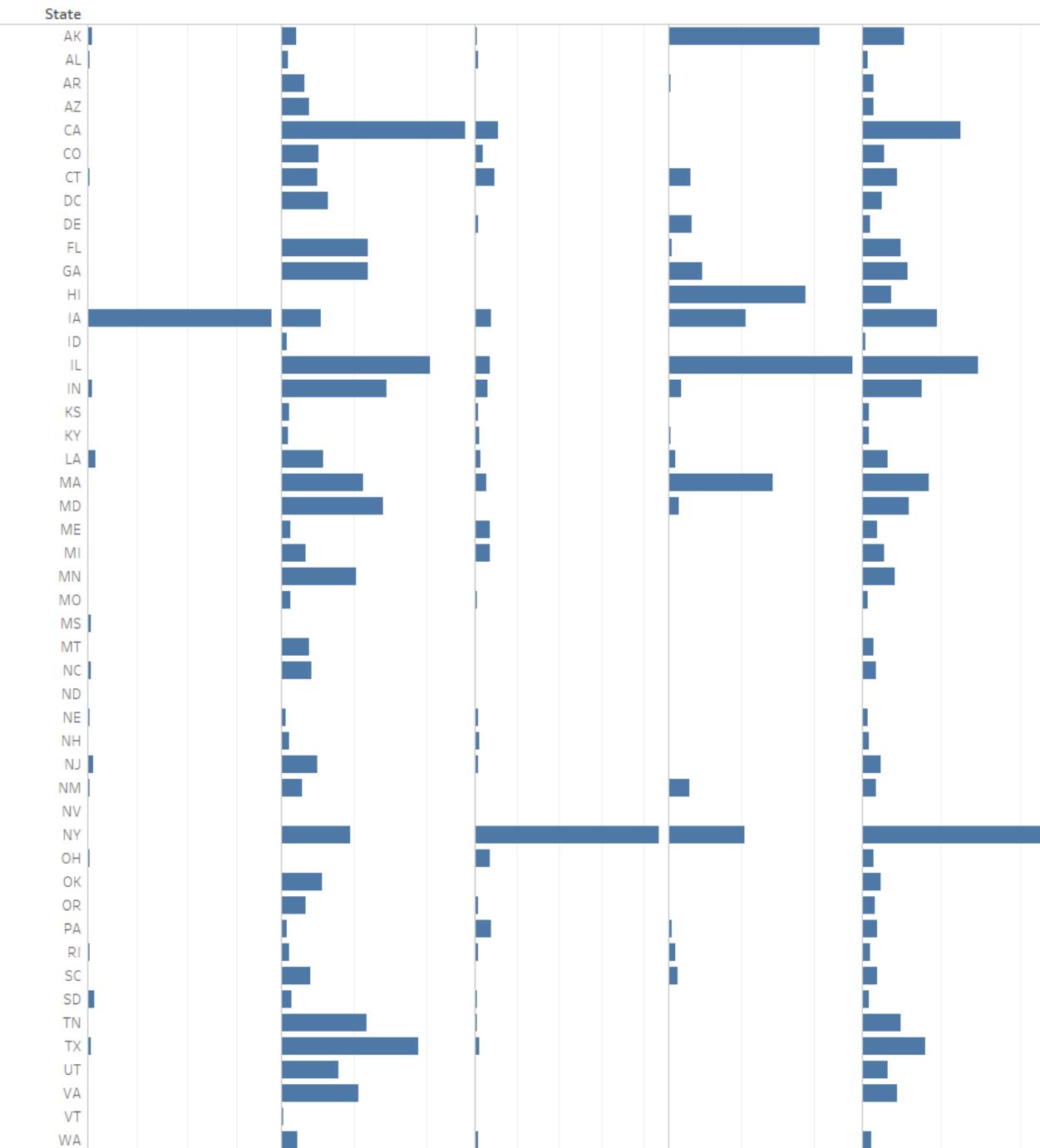


Population for each State.

State, sum of Librarians and sum of MLS Librarians. Color shows sum of MLS Librarians. Size shows sum of Librarians. The marks are labeled by State, sum of Librarians and sum of MLS Librarians.

Data Exploration

Revenues

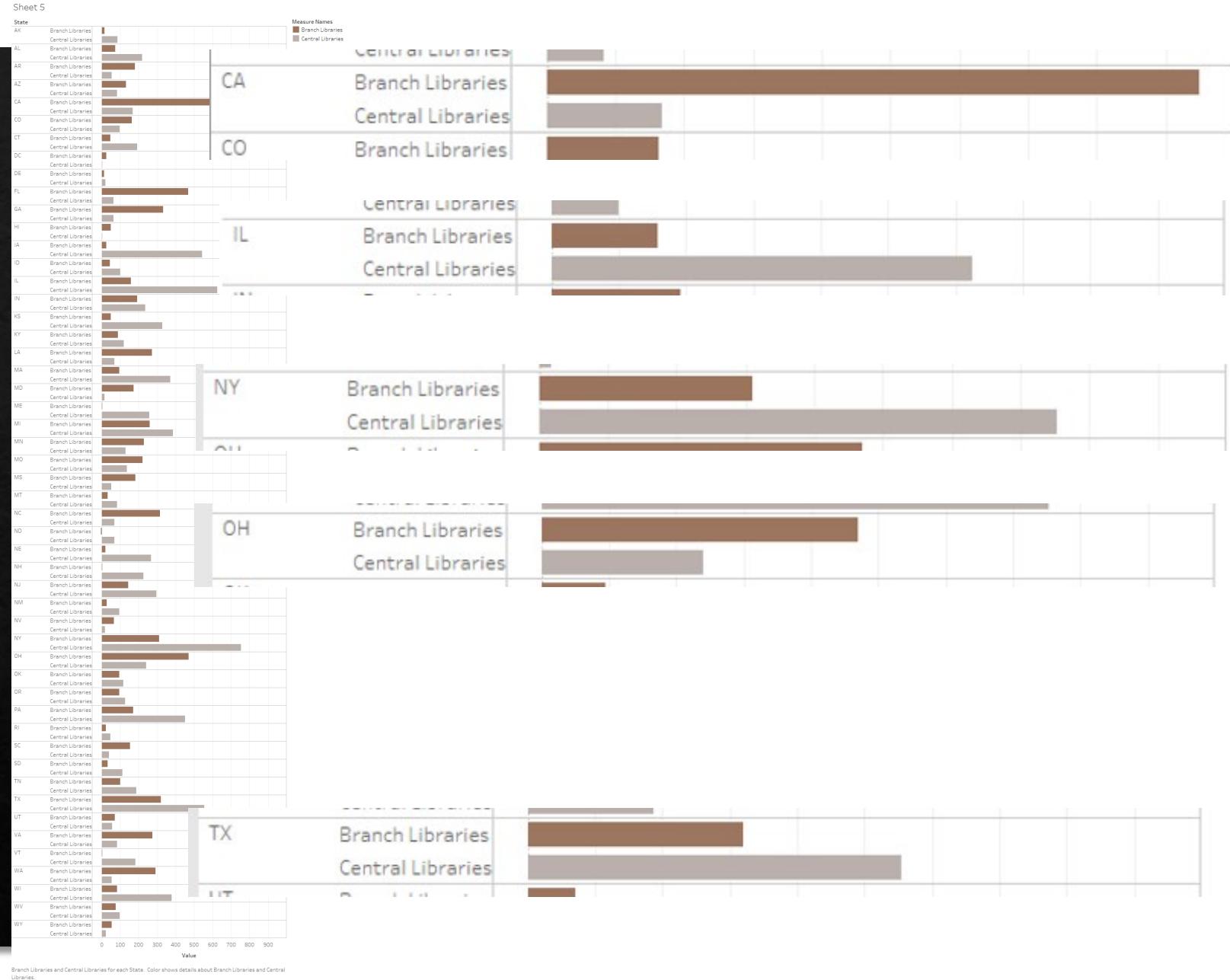


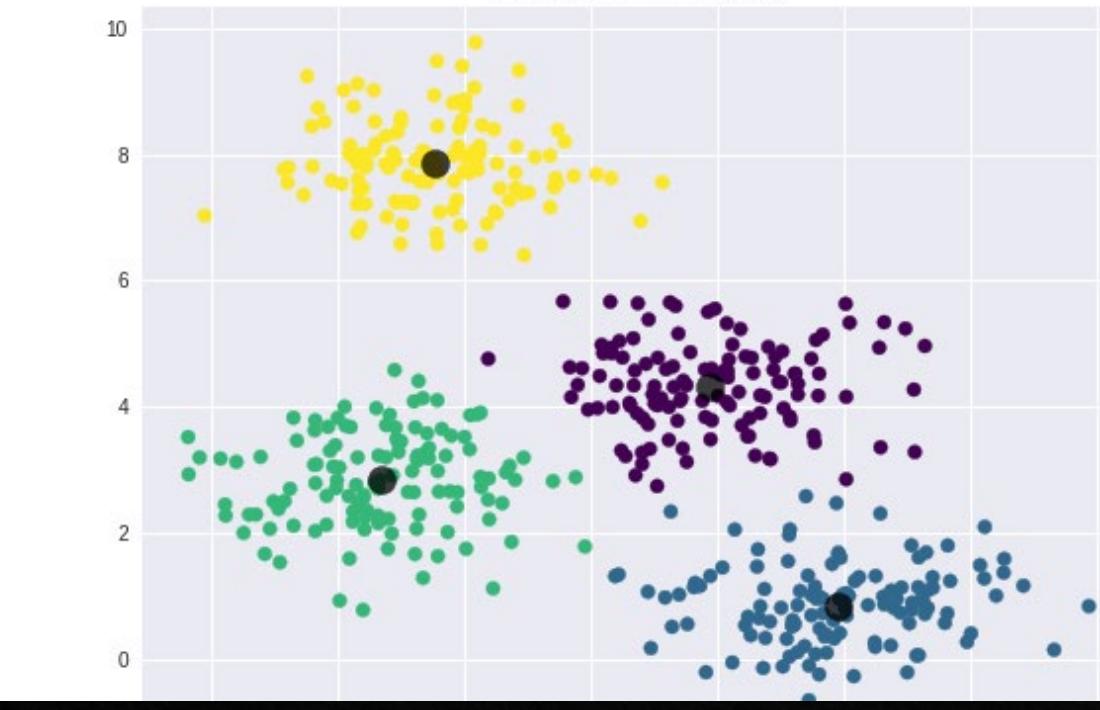
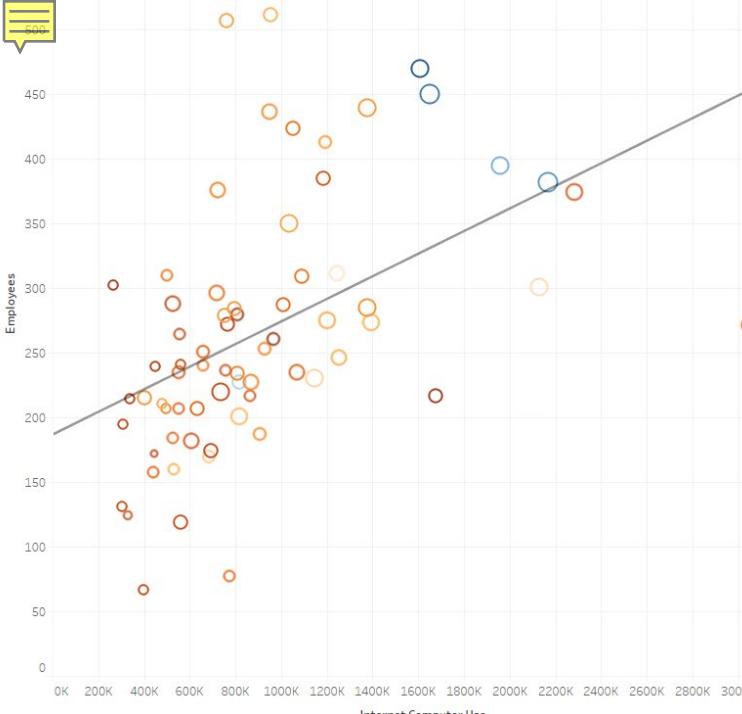
RevenueVsExpenditure



Data Exploration

- ❖ Branch Libraries and Central Libraries per state
 - ❖ CA
 - ❖ NY
 - ❖ TX
 - ❖ IL
 - ❖ OH





Approach & Techniques

- ❖ Challenges faced
 - ❖ Difficulty in interpreting data as there were huge variations in most attributes like library programs, revenue, staff, expenditure etc. Hence the team decided to cluster data and then research to get accurate results
- ❖ Models
 - ❖ Clustering
 - ❖ Linear Regression

K-Means

- ❖ 4 Clusters based on
 - ❖ Population
 - ❖ Revenue
 - ❖ Staff
 - ❖ Number of Central Libraries

	Absolute	Percent
Within-class	6620672676052.600	18.26%
Between-classes	29636671836926.500	81.74%
Total	36257344512979.100	100.00%

Class centroids:								
Class	County Population	Service Population Without Duplic	Central Libraries	Total Staff	Total Operating Revenue	Sum of weights	Within-class variance	
1	288388.334	16567.803	0.988	7.424	552923.003	8765.000	1326374359992.020	
2	1186918.897	208808.023	0.861	96.615	8814354.175	388.000	20981303924074.800	
3	1258279.704	853238.718	0.704	377.614	37974968.225	71.000	178632591150906.000	
4	4373992.875	2537876.500	0.750	1143.215	138546669.875	8.000	4121025223494720.000	

Class	County Population	Service Population	Central Libraries	Total Staff	Total Operating Revenue	# of braries	ClusterName based on Service Population
1	288388.334	16567.803	0.988	7.424	552923.003	8765.000	Small - Comparitively small target audience, lesser operating revenue but represents 94% of US population. It is seen that they have fewer staff to manage libraries
2	1186918.897	208808.023	0.861	96.615	8814354.175	388.000	Medium - This cluster has medium Operating revenue and staffed at an average of 96 and represents 4% of US library population
3	1258279.704	853238.718	0.704	377.614	37974968.225	71.000	Large - Well staffed and has an operating revenue of \$37million. Represents ~1% of US library population
4	4373992.875	2537876.500	0.750	1143.215	138546669.875	8.000	XL - Superlative numbers - high staffing,\$138million operating revenue but it represents less than 0.1% of the US library population

Linear Regression	Small	Medium
Intercept	3180.25	-43464.94
Service Population Without Dups	0.44	0.46
Total Staff	2743.35	1499.10
Total Operating Revenue	0.02	0.03
Registered Users	1.10	5.23
Library Programs	20.53	2.00
Internet Computer Use	0.52	27.44
Wireless Internet Sessions	0.01	0.50
Hours Open	-1.06	0.02
Total Collection Expenditures	0.20	0.05
R- Squared	81%	83%

Linear Regression

Linear Regression	Large
Intercept	812998.69
Total Staff	1966.56
Video	2.55
Internet Computer Use	1.036
Library Programs	37.11
Circulation Transactions	0.087
Adjusted R-Squared	68%

Recommendations

- ❖ Regardless of the size of the library, these four features persisted as the most important features to draw customers.

