

CS 4460 P5 Report

Team Members: Yingshan CHANG, Aarti Thapar

Dataset chosen: colleges.csv

In colleges.csv, our visualization contains the following columns that we used in our scatterplot matrix and bar chart: Name, Control, Region, Locale, White, Black, Hispanic, Asian, Expenditure, MedianDebt and MedianFamilyIncome.

Name	Predominant	Highest	Deg	Control	Region	Locale	Admission R	ACT Median	SAT Average	Undergrad P	White	Black	Hispanic	Asian	AmericanInd	PacificIsland	Biracial	Nonresident	% Part-time	Average	
Michigan Jewish Institute	3	3	Private	Great Lakes	Large Suburb	0.981	0	0	1642	0.9888	0	0	0	0	0	0	0	0.0012	0.2314	15	
College of the Ozarks	3	3	Private	Great Plains	Distant Town	0.1301	23	1050	1513	0.9174	0.0059	0.0198	0.0053	0.0059	0.0059	0	0.0172	0.0218	0.0099	26	
University of Puerto Rico-Arecib	3	3	Public	Outing Are Small City	0.5302	0	0	3654	0	0	1	0	0	0	0	0	0	0.0613	0	11	
University of Puerto Rico-Bayar	3	3	Public	Outing Are Large Suburb	0.2547	0	0	4847	0	0	1	0	0	0	0	0	0	0.1192	0	8	
University of Puerto Rico-Carolin	3	3	Public	Outing Are Mid-size City	0.655	0	1029	3880	0.0077	0.0026	0.9126	0	0.0015	0	0	0.0015	0	0.0351	0	11	
University of Puerto Rico-Cayey	3	3	Public	Outing Are Large Suburb	0.8342	0	1094	3779	0	0	1	0	0	0	0	0	0	0.072	0	11	
University of Puerto Rico-Humau	3	3	Public	Outing Are Large Suburb	0.4487	0	0	3383	0.0059	0.0095	0.8835	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0213	11
University of Puerto Rico-Agud	3	3	Public	Outing Are Large Suburb	0.9234	0	0	1038	2686	0.0073	0.001	0.9106	0.0003	0.0003	0	0.0017	0	0.0655	0	8	
University of Puerto Rico-Ponce	3	3	Public	Outing Are Mid-size City	0.6398	0	0	3120	0	0	1	0	0	0	0	0	0	0	0.0766	13	
Berea College	3	3	Private	Great Plains	Distant Town	0.3401	24	1113	1597	0.6446	0.1512	0.0423	0.0154	0.0035	0.0019	0.051	0.0756	0.0039	0	20	
Baylor Central University	3	4	Private	Outing Are Large Suburb	0.7368	0	0	1654	0	0.006	0.994	0	0	0	0	0	0	0	0.2213	11	
University of Puerto Rico-Mayag	3	4	Public	Outing Are Small City	0.725	0	0	10697	0	0	1	0	0	0	0	0	0	0	0.0526	14	
University of Puerto Rico-Pri	3	4	Public	Outing Are Large City	0.5248	0	1195	11834	0.007	0.0014	0.9027	0.0014	0.001	0.0001	0.0001	0.0001	0.0001	0.0001	0.1032	8	
Brigham Young University-Hawi	3	4	Private	Far West	Fringe Town	0.3604	24	1081	2664	0.3566	0.0034	0.0556	0.0396	0.0023	0.0706	0.0867	0.3885	0.0661	14		
Duke University	3	4	Private	Southeast	Mid-size City	0.1342	32	1444	6501	0.4861	0.1001	0.0611	0.2124	0.0062	0.0005	0.0174	0.0892	0.0009	59		

Expenditure	Average	Faci	% Full-time	% Undergrad	Completion	Retention	R2	% Undergrad	3 Year	Defau	MedianDebt	Median Debt	Median Deft	FederalLoan	% Pell Grant	Average Age	Average Fan	MedianFam	Poverty Rate	Number of
10970	3433	0.333	0.796	0.175	0.441	0.2828	6.6	0	0	0.02	1	19.510000	2657.5	27033	14.6959998	12				
9447	6793	0.6471	0.2166	0.6454	0.8691	0.0205	0	0	0	0	0	0	0	14.0369997	33369.07	32674.5	10.1400003	3		
6585	5133	0.9158	0.8188	0.4004	0.7941	0.0575	11.9	4500	4600	4500	0.33000001	0.98000002	15.5499992	25464.65	24433.5	45.7299995				
4930	5738	0.7742	0.6527	0.3924	0.856	0.0802	11.9	4500	5500	4500	0.33000001	0.98000002	15.5499992	25464.65	24433.5	45.7299995				
6675	5828	0.5385	0.6995	0.3327	0.8367	0.0746	11.8	4500	4957.5	4500	0.33000001	0.98000002	15.5499992	25464.65	24433.5	45.7299995				
5905	7195	0.8431	0.7385	0.4638	0.676	0.0419	11.8	4500	5000	4000	0.33000001	0.98000002	15.5499992	25464.65	24433.5	45.7299995				
9908	5812	0.8464	0.7527	0.4343	0.868	0.0658	11.8	4500	5000	3800	0.33000001	0.98000002	15.5499992	25464.65	24433.5	45.7299995				
5668	5170	0.7697	0.811	0.321	0.831	0.0696	11.8	5000	5500	4500	0.33000001	0.98000002	15.5499992	25464.65	24433.5	45.7299995				
5276	5174	0.7072	0.8067	0.4072	0.8673	0.0756	11.9	5250	5500	4500	0.33000001	0.98000002	15.5499992	25464.65	24433.5	45.7299995				
15067	7357	0.8444	0.8293	0.6167	0.8107	0.0746	12.8	5496.5	5750	4676.5	0.55000001	0.97000003	20.0499992	30776.75	29786.5	12.2299995	5			
2704	2630	1	0.7144	0.2752	0.7017	0.2849	16.4	5500	7500	5325	0.30000001	0.98000002	23.4699993	16277.49	12591.40	3400002	46			
8372	6533	0.9807	0.6081	0.4284	0.8932	0.0546	11.6	5500	5500	5500	0.33000001	0.98000002	15.5499992	25464.65	24433.5	45.7299995				
12517	7104	0.7862	0.5895	0.4748	0.8966	0.0746	11.9	5500	5500	5197	0.33000001	0.98000002	15.5499992	25464.65	24433.5	45.7299995				
11999	8794	1	0.2985	0.4703	0.6132	0.224	6.9	5780	10000	5500	0.64999998	0.82999998	22.2399998	39733.27	22339	8.65999985	14			
51653	14288	0.94	0.1366	0.9423	0.972	0.0038	1.1	6569	6500	7000	0.95999998	0.24	21.3099995	92670.02	73265	6.34999991	10			
8154	7001	0.3626	0.5399	0.2635	0.7415	0.2812	10.1	7000	10000	5500	0.23	0.93000001	24.1800003	25815.23	21863	16.6000004	27			

Figure a: Our data set in colleges.csv. Columns highlighted are the data points we used in our visualization.

With the help of our visualization, we wanted our user to be able to do a couple of things that would aid their understanding about our dataset. We wanted them to:

- Be able to visualize the data by locality, regions and controls (i.e whether the university is a public or private university).
- Be able to visualize the relationship between the Median Debt, Expenditure and Median Family Income of university student families, in terms of the scatterplot matrix.
- Be able to visualize the ethnicity distribution between different regions, localities and controls.
- Be able to view information about finances and ethncical distributions simultaneously and showing a link between them by using brushing and linking.

In terms of the design of our visualization, our main aim was to create a design that can communicate to the viewer a relationship between financial and ethnic distribution of college students in a broader scale. However, there are also other questions that our visualization aims to answer to the viewer such as:

- Is there a certain type of ethnicity that is more prominent in the data set of colleges that contains data about college students?
- Is there a relationship between family debt, expenditure and family income? If so, is it positively or negatively correlated or random? This is extremely useful in understanding

CS 4460 P5 Report

Team Members: Yingshan CHANG, Aarti Thapar

the financial status of students and their families and gives us data about what financial status of students enroll in which universities in the United States.

- Are prominent ethnicities different between each region and locality?
- What are the maximum and minimum values of correlation between family debt, expenditure and family income and what are the names of the universities that are in these data points? We can find this out by hovering over the data point as the visualization gives information about the data point such as university name and locality/region/Control depending on what filter is selected.

Moreover, we made sure to improve our visualization by using certain design concepts that we learned in class to aid our viewer in understanding the data set. We decided to use brushing and linking as our concept of designing an interactive visualization as it defines the domain of our attention and we can visualize the distribution of ethnicity and finances in a particular domain in the scatterplot matrix, and thus get that information also visualized in the grouped bar chart through linking.

In terms of our brushing functionality, when the user selects a region in our scatterplot matrix, the data points within the region are shown and the data points outside the domain of the region are hidden. Next, the linking functionality is used as after the user selects a region, the user can also select a filter button from the three options in the bottom left, then the corresponding data points are shown on the grouped bar chart (with percentages of ethnic backgrounds shown as well). The user also has the freedom to drag and resize their brush window and the data in the bar chart also changes in real-time.

In addition, the user also gets information about what percentage of ethnic backgrounds are present in each area corresponding to the data in the scatterplot matrix in the grouped bar chart; this can be done if they hover on each column of the grouped bar chart and a small display window is shown with this information. Moreover, this visualization also includes two legends for both the scatterplot matrix and grouped bar chart, which will help the user to identify what the data points are in the charts.

Here are some screenshots of the features of our visualization:

CS 4460 P5 Report

Team Members: Yingshan CHANG, Aarti Thapar

Colleges in US: Financial Statuses & Ethnic Distributions

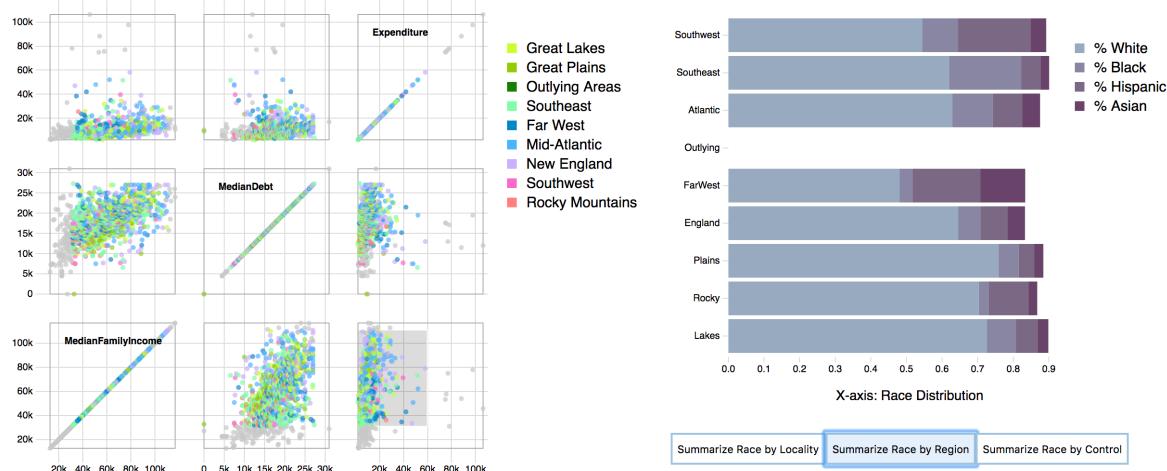


Figure 1: Visualization of scatterplot matrix and grouped bar chart filtered by Locality with their legends.

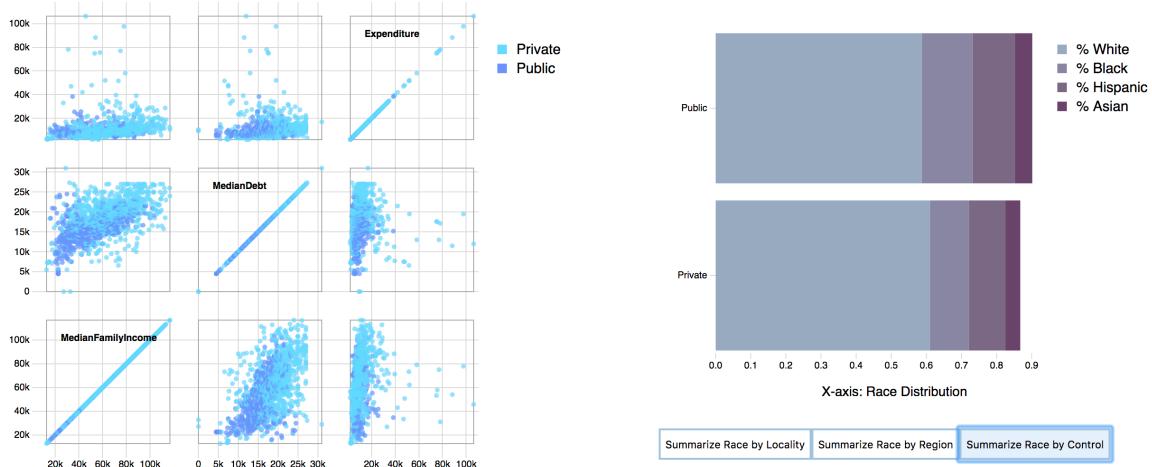


Figure 2: Visualization of scatterplot matrix and grouped bar chart filtered by Control (Private or Public).

CS 4460 P5 Report

Team Members: Yingshan CHANG, Aarti Thapar



Figure 3: Visualization of grouped bar chart hover functionality

Figure 4: Visualization of scatterplot matrix chart hover functionality

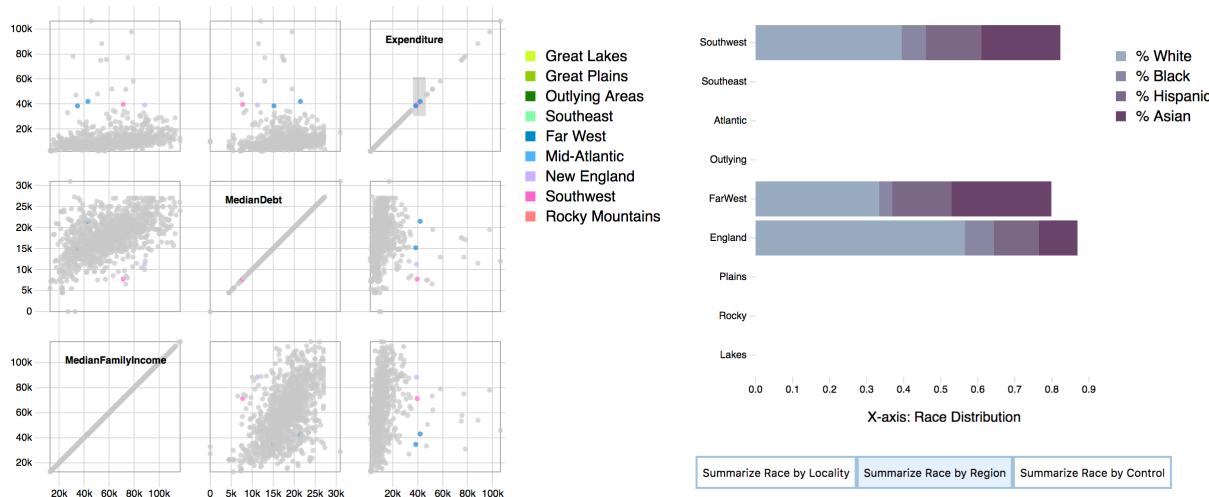


Figure 5: Visualization of scatterplot matrix and grouped bar chart after brushing is performed on a region is selected via the brush on the scatterplot matrix. The corresponding data points are shown in the grouped bar chart in real-time.