



Collecting Storing & Retrieving Data

SPRING 2017

HDI Analysis across Cities in USA

By, Harshitha Krishne Gowda, Ashwin Rajkumar



INDEX

Table of contents:	Page
• Overview	3
○ Problems Driving The Project	3
• Scraping The Data	4
○ Processes Implemented	4
○ Technical Choices	4
• Issues	5
• Inputs	6
• Outputs	6
• Conclusion	13

OVERVIEW

The Human Development Index (HDI) is a composite statistic of life expectancy, education, and per capita income indicators, which are used to rank countries into four tiers of human development. A country scores higher HDI when the lifespan is higher, the education level is higher, and the GDP per capita is higher. All data scraped in the present project is from the 2009-2013 American Community Survey 5-Year Estimates. Here in this project we would collect the data like Per Capita Income, Median household Income, Median family income, Population and Number of households for each city and import into a data frame in R. The data would then be cleaned to get the data that would be used for analysis.

PROBLEMS DRIVING THE PROJECT

- Finding the city with highest Per Capita Income, Median household Income, Median family income for each state and then comparing each state
- Finding the top 50 cities with highest Per Capita Income, Median household Income, Median family income and concluding which state has most cities to interpret the best state
- Finding the average Per Capita Income, Median household Income, Median family income for each state and then analysing visually by graphs
- Finding the total population of each state from the collected data
- Obtaining bar graphs for the above listed problems

SCRAPING THE DATA

We used packages like Rcurl and XML to scrape the required data programmatically from Wikipedia page – ‘List of United States counties by per capita income’ and transformed the content into analysable form. The unwanted data in the dataset was removed and then after cleaning the data it was stored in a data frame.

PROCESSES IMPLEMENTED

- Used Packages like Rcurl, XML, RSQLite, dplyr, ggplot2
- Cleaned the data and stored it in a dataframe
- Performed operations like Average, Greatest, Summation
- Compared the data from the obtained analysis
- Obtained bar graphs for states with highest Per Capita Income, Median household Income, Median family income
- Stored it in a SQLite database storage system

TECHNICAL CHOICES

- Rcurl: The Rcurl package is an R-interface to the libcurl library that provides HTTP facilities
This allows us to download files from Web servers by GETting forms
The primary top-level entry points are: `getURL()`, `getURLContent()`
- XML: The XML package is necessary to parse the XML and HTML code
This also offers access to an XPath "interpreter"
- RSQLite: SQLite is an open source, embedded relational database which helps to directly import a data frame into a database connection

Allows us to use queries to communicate with the database

- Dplyr: It provides simple verbs, functions that correspond to the most common data manipulation tasks, to help translate those thoughts into code

It uses efficient data storage back ends

- Barplot: It is used to show visually the results obtained after analysis
- We used for loops to iterate through each level and then called the function repeatedly

ISSUES

- Issue: Scraping the data from Wikipedia was quite difficult as we found it hard to scrape using X-path and tools like selector gadget

Solution: We made use of readhtml table and scraped the required data from the webpage

- Issue: The data to be scraped was in multiple tables and it was difficult to scrape it altogether into a single data frame

Solution: We scraped each table individually and stored in separate data frames and finally concatenated into a single data frame

- Issue: The scraped data was not in required numeric format and had dollar signs and commas along with unwanted rows

Solution: We made use of gsub function to remove the unwanted characters and converted the mode into numeric

- Issue: As we had a lot of variables, all the labels in the x-axis were not visible

INPUT

```
1 install.packages("RCurl")
2 library(RCurl)
3 install.packages("XML")
4 library(XML)
5
6 # Getting URL data from the website specified in the getURL function.
7 webpage <- getURL("https://en.wikipedia.org/wiki/List_of_United_States_counties_by_per_capita_income") # gets the URL
8 # Collecting the data from the tables in the web page and storing in data frames.
9 table1 <- readHTMLTable(webpage, which = 2)
10 table2 <- readHTMLTable(webpage, which = 3)
11 table3 <- readHTMLTable(webpage, which = 4)
12 table4 <- readHTMLTable(webpage, which = 5)
13 table5 <- readHTMLTable(webpage, which = 6)
14 table6 <- readHTMLTable(webpage, which = 7)
15 table7 <- readHTMLTable(webpage, which = 8)
16 table8 <- readHTMLTable(webpage, which = 9)
17 table9 <- readHTMLTable(webpage, which = 10)
18 table10 <- readHTMLTable(webpage, which = 11)
19
20 # Combining all the individual data frames into a single one using rbind.
21 HDI <- rbind(table1, table2, table3, table4, table5, table6, table7, table8, table9, table10)
22 nrow(HDI)
23
24 # Few rows had some inappropriate data with the word State and Country under the State column. So these were removed.
25 ind <- which(HDI$State == "State")
26 ind
27 HDI <- HDI[-ind,]
28 ind1 <- which(HDI$State == "Country")
29 ind1
30 HDI <- HDI[-ind1,]
31
32 # changing the names of columns
33 colnames(HDI) <- c("Rank", "County-equivalent", "State", "Per_capita_income", "Median_household_income", "Median_family_income")
34 HDI
35 class(HDI$Per_capita_income)
36 <
```

OUTPUTS

```
The city Shelby has the highest median household income of $68770 for the state Alabama
The city Aleutians West Census Area has the highest median household income of $81853 for the state Alaska
The city Santa Clara has the highest median household income of $91702 for the state California
The city Douglas has the highest median household income of $101591 for the state Colorado
The city Fairfield has the highest median household income of $82283 for the state Connecticut
The city New Castle has the highest median household income of $64537 for the state Delaware
The city Washington City has the highest median household income of $65830 for the state District of Columbia
The city St. Johns has the highest median household income of $64876 for the state Florida
The city Forsyth has the highest median household income of $86569 for the state Georgia
The city Honolulu has the highest median household income of $72764 for the state Hawaii
The city Blaine has the highest median household income of $64042 for the state Idaho
The city Kendall has the highest median household income of $81765 for the state Illinois
The city Hamilton has the highest median household income of $82468 for the state Indiana
The city Dallas has the highest median household income of $73847 for the state Iowa
The city Johnson has the highest median household income of $74717 for the state Kansas
The city Oldham has the highest median household income of $83391 for the state Kentucky
The city Ascension has the highest median household income of $69070 for the state Louisiana
The city Cumberland has the highest median household income of $57461 for the state Maine
The city Howard has the highest median household income of $109865 for the state Maryland
The city Nantucket has the highest median household income of $85478 for the state Massachusetts
The city Livingston has the highest median household income of $72359 for the state Michigan
The city Scott has the highest median household income of $86112 for the state Minnesota
The city Madison has the highest median household income of $59904 for the state Mississippi
The city St. Charles has the highest median household income of $71077 for the state Missouri
The city Jefferson has the highest median household income of $60863 for the state Montana
The city Sarpy has the highest median household income of $69965 for the state Nebraska
The city Lander has the highest median household income of $72742 for the state Nevada
```

Highest median household income for each state.

The city Rockingham has the highest median household income of \$77348 for the state New Hampshire
 The city Hunterdon has the highest median household income of \$106143 for the state New Jersey
 The city Los Alamos has the highest median household income of \$106686 for the state New Mexico
 The city Nassau has the highest median household income of \$97690 for the state New York
 The city Wake has the highest median household income of \$66006 for the state North Carolina
 The city Williams has the highest median household income of \$76210 for the state North Dakota
 The city Delaware County has the highest median household income of \$89757 for the state Ohio
 The city Clackamas has the highest median household income of \$64352 for the state Oregon
 The city Chester has the highest median household income of \$86050 for the state Pennsylvania
 The city Washington County has the highest median household income of \$72138 for the state Rhode Island
 The city Beaufort has the highest median household income of \$57316 for the state South Carolina
 The city Lincoln has the highest median household income of \$74751 for the state South Dakota
 The city has the highest median household income of \$ for the state
 The city Williamson has the highest median household income of \$89779 for the state Tennessee
 The city Rockwall has the highest median household income of \$86119 for the state Texas
 The city Summit has the highest median household income of \$83336 for the state Utah
 The city Chittenden has the highest median household income of \$63989 for the state Vermont
 The city Loudoun has the highest median household income of \$122238 for the state Virginia
 The city King has the highest median household income of \$71811 for the state Washington
 The city Waukesha has the highest median household income of \$75850 for the state Wisconsin
 The city Campbell has the highest median household income of \$79488 for the state Wyoming
 The city Maricopa has the highest median household income of \$53596 for the state Arizona
 The city Saline has the highest median household income of \$55348 for the state Arkansas
 The city has the highest median household income of \$ for the state
 The city Canadian has the highest median household income of \$63629 for the state Oklahoma
 The city Jefferson has the highest median household income of \$65304 for the state West Virginia
 >

Highest median household income for each state.(cont)

The city Shelby has the highest median family income of \$82558 for the state Alabama
 The city Denali has the highest median family income of \$99091 for the state Alaska
 The city Marin has the highest median family income of \$117357 for the state California
 The city Douglas has the highest median family income of \$112930 for the state Colorado
 The city Fairfield has the highest median family income of \$104250 for the state Connecticut
 The city New Castle has the highest median family income of \$79954 for the state Delaware
 The city Washington City has the highest median family income of \$79665 for the state District of Columbia
 The city St. Johns has the highest median family income of \$80707 for the state Florida
 The city Forsyth has the highest median family income of \$99573 for the state Georgia
 The city Kalawao has the highest median family income of \$88750 for the state Hawaii
 The city Blaine has the highest median family income of \$76575 for the state Idaho
 The city DuPage has the highest median family income of \$95208 for the state Illinois
 The city Hamilton has the highest median family income of \$99768 for the state Indiana
 The city Dallas has the highest median family income of \$90092 for the state Iowa
 The city Johnson has the highest median family income of \$92512 for the state Kansas
 The city Oldham has the highest median family income of \$91587 for the state Kentucky
 The city Ascension has the highest median family income of \$77640 for the state Louisiana
 The city Cumberland has the highest median family income of \$75432 for the state Maine
 The city Howard has the highest median family income of \$128657 for the state Maryland
 The city Norfolk has the highest median family income of \$108943 for the state Massachusetts
 The city Washtenaw has the highest median family income of \$85115 for the state Michigan
 The city Scott has the highest median family income of \$97693 for the state Minnesota
 The city Madison has the highest median family income of \$73775 for the state Mississippi
 The city St. Charles has the highest median family income of \$83907 for the state Missouri
 The city Jefferson has the highest median family income of \$76469 for the state Montana
 The city Washington County has the highest median family income of \$83670 for the state Nebraska
 The city Eureka has the highest median family income of \$94648 for the state Nevada

Highest median family income for each state.

The city Rockingham has the highest median family income of \$93420 for the state New Hampshire
 The city Hunterdon has the highest median family income of \$125828 for the state New Jersey
 The city Los Alamos has the highest median family income of \$124979 for the state New Mexico
 The city Nassau has the highest median family income of \$112874 for the state New York
 The city Orange has the highest median family income of \$84770 for the state North Carolina
 The city Williams has the highest median family income of \$86285 for the state North Dakota
 The city Delaware County has the highest median family income of \$103564 for the state Ohio
 The city Washington County has the highest median family income of \$76711 for the state Oregon
 The city Chester has the highest median family income of \$105278 for the state Pennsylvania
 The city Bristol has the highest median family income of \$93131 for the state Rhode Island
 The city Beaufort has the highest median family income of \$66878 for the state South Carolina
 The city Lincoln has the highest median family income of \$85666 for the state South Dakota
 The city has the highest median family income of \$ for the state
 The city Williamson has the highest median family income of \$104796 for the state Tennessee
 The city Glasscock has the highest median family income of \$98409 for the state Texas
 The city Summit has the highest median family income of \$95879 for the state Utah
 The city Chittenden has the highest median family income of \$83887 for the state Vermont
 The city Falls Church City has the highest median family income of \$152857 for the state Virginia
 The city King has the highest median family income of \$92510 for the state Washington
 The city Waukesha has the highest median family income of \$92974 for the state Wisconsin
 The city Sublette has the highest median family income of \$85885 for the state Wyoming
 The city Maricopa has the highest median family income of \$63910 for the state Arizona
 The city Saline has the highest median family income of \$66936 for the state Arkansas
 The city has the highest median family income of \$ for the state
 The city Canadian has the highest median family income of \$72849 for the state Oklahoma
 The city Jefferson has the highest median family income of \$77447 for the state West Virginia

Highest median family income for each state.(cont)

The city Shelby has the highest per_capita_income of \$33313 for the state Alabama
 The city North Slope has the highest per_capita_income of \$46457 for the state Alaska
 The city Marin has the highest per_capita_income of \$56791 for the state California
 The city Pitkin has the highest per_capita_income of \$51814 for the state Colorado
 The city Fairfield has the highest per_capita_income of \$48721 for the state Connecticut
 The city New Castle has the highest per_capita_income of \$32406 for the state Delaware
 The city Washington City has the highest per_capita_income of \$45290 for the state District of Columbia
 The city St. Johns has the highest per_capita_income of \$36836 for the state Florida
 The city Fulton has the highest per_capita_income of \$36757 for the state Georgia
 The city Kalawao has the highest per_capita_income of \$45515 for the state Hawaii
 The city Blaine has the highest per_capita_income of \$35524 for the state Idaho
 The city DuPage has the highest per_capita_income of \$38570 for the state Illinois
 The city Hamilton has the highest per_capita_income of \$39521 for the state Indiana
 The city Dallas has the highest per_capita_income of \$35844 for the state Iowa
 The city Johnson has the highest per_capita_income of \$38827 for the state Kansas
 The city Oldham has the highest per_capita_income of \$33591 for the state Kentucky
 The city St. Tammany has the highest per_capita_income of \$30363 for the state Louisiana
 The city Cumberland has the highest per_capita_income of \$32880 for the state Maine
 The city Montgomery has the highest per_capita_income of \$49038 for the state Maryland
 The city Nantucket has the highest per_capita_income of \$47331 for the state Massachusetts
 The city Oakland has the highest per_capita_income of \$36458 for the state Michigan
 The city Hennepin has the highest per_capita_income of \$37485 for the state Minnesota
 The city Madison has the highest per_capita_income of \$33170 for the state Mississippi
 The city Platte has the highest per_capita_income of \$35143 for the state Missouri
 The city Madison has the highest per_capita_income of \$33062 for the state Montana
 The city Sioux has the highest per_capita_income of \$31517 for the state Nebraska
 The city Douglas has the highest per_capita_income of \$34123 for the state Nevada

Highest per capita income for each state.

The city Rockingham has the highest per_capita_income of \$38399 for the state New Hampshire
 The city Hunterdon has the highest per_capita_income of \$50349 for the state New Jersey
 The city Los Alamos has the highest per_capita_income of \$51044 for the state New Mexico
 The city New York County has the highest per_capita_income of \$62498 for the state New York
 The city Orange has the highest per_capita_income of \$34465 for the state North Carolina
 The city Williams has the highest per_capita_income of \$38738 for the state North Dakota
 The city Delaware County has the highest per_capita_income of \$40189 for the state Ohio
 The city Clackamas has the highest per_capita_income of \$32781 for the state Oregon
 The city Chester has the highest per_capita_income of \$42210 for the state Pennsylvania
 The city Newport has the highest per_capita_income of \$40293 for the state Rhode Island
 The city Beaufort has the highest per_capita_income of \$32503 for the state South Carolina
 The city Union has the highest per_capita_income of \$36167 for the state South Dakota
 The city has the highest per_capita_income of \$ for the state
 The city Williamson has the highest per_capita_income of \$41292 for the state Tennessee
 The city Borden has the highest per_capita_income of \$50042 for the state Texas
 The city Summit has the highest per_capita_income of \$42462 for the state Utah
 The city Chittenden has the highest per_capita_income of \$33281 for the state Vermont
 The city Arlington has the highest per_capita_income of \$62018 for the state Virginia
 The city King has the highest per_capita_income of \$39911 for the state Washington
 The city Ozaukee has the highest per_capita_income of \$42041 for the state Wisconsin
 The city Teton has the highest per_capita_income of \$43444 for the state Wyoming
 The city Maricopa has the highest per_capita_income of \$27256 for the state Arizona
 The city Pulaski has the highest per_capita_income of \$27414 for the state Arkansas
 The city has the highest per_capita_income of \$ for the state
 The city Tulsa has the highest per_capita_income of \$27676 for the state Oklahoma
 The city Jefferson has the highest per_capita_income of \$29605 for the state West Virginia

Highest per capita income for each state.(cont)

Rank	County-equivalent	State	Per_capita_income	Median_household_income	Median_family_income	Population	Number_of_households
1	New York County	New York	62498	69639	84627	1605272	736192
2	Arlington	Virginia	62018	103200	139244	214061	94454
3	Falls Church City	Virginia	59088	120000	152857	12731	5020
4	Marin	California	56791	90839	117357	254643	102912
5	Alexandria City	Virginia	54608	85706	107511	143684	65369
6	Pitkin	Colorado	53924	72745	93981	17173	7507
7	Los Alamos	New Mexico	51044	106686	124979	17979	7590
8	Fairfax County	Virginia	50532	110292	128596	1101071	389908
9	Hunterdon	New Jersey	50349	106143	125828	127847	46816
10	Borden	Texas	50042	71607	76258	849	225
11	Montgomery	Maryland	49838	98221	117408	989474	368563
12	Morris	New Jersey	48814	98633	117683	495261	179665
13	Fairfield	Connecticut	48721	82283	104250	926233	332655
14	San Francisco	California	48486	75604	98116	817501	345344
15	Howard	Maryland	48172	100805	128657	203821	106142
16	Westchester	New York	47984	81946	105341	956283	343561
17	Somerset	New Jersey	47803	99020	117393	326207	115531
18	Nantucket	Massachusetts	47331	85478	92500	10224	4069
19	Leidoun	Virginia	46565	122238	133550	326477	106997
20	North Slope	Alaska	46457	80761	82500	9484	1090
21	San Mateo	California	45732	88202	105844	729543	257941
22	Kalawao	Hawaii	45515	59375	88750	71	46
23	Washington City	District of Columbia	45290	65830	79665	610371	263649
24	Goochland	Virginia	45039	80976	92842	21565	8050
25	Norfolk	Massachusetts	44692	84916	108943	677296	257451
26	Fairfax City	Virginia	44345	97242	113538	23027	8492
27	Douglas	Colorado	43634	101591	112930	293814	103700
28	Teton	Wyoming	43444	68078	83558	21575	7503
29	Bergen	New Jersey	43347	83704	102222	912795	335422
30	Middlesex	Massachusetts	42861	82090	104832	1522533	581120
31	Monmouth	New Jersey	42749	84526	105971	629735	233909
32	Summit	Utah	42462	83336	95879	37232	13627
33	Massau	New York	42400	97690	112074	1343705	441955
34	Mineral	Colorado	42255	47083	68438	761	393
35	Chester	Pennsylvania	42210	86050	105278	503075	184788
36	Ozaukee	Wisconsin	42041	75457	92262	86708	34108
37	Clear Creek	Colorado	41716	67259	89583	9048	4024
38	Santa Clara	California	41513	91702	103983	1812208	609377
39	Montgomery	Pennsylvania	41472	79183	99587	804621	307488
40	Williamson	Tennessee	41292	89779	104796	188935	66364
41	San Riquel	Colorado	40827	63594	77708	7496	3234
42	Anne Arundel	Maryland	40415	87430	101268	544426	199004
43	Putnam	New York	40309	95117	109108	99718	34496
44	Newport	Rhode Island	40293	71713	88962	82545	34774
45	Delaware County	Ohio	40189	89757	103564	178139	64208
46	Middlesex	Connecticut	39992	70994	88081	165090	66141
47	King	Washington	39911	71811	92510	1974567	802686
48	Fauquier	Virginia	39608	88409	103556	66015	23130
49	Hamilton	Indiana	39521	82468	99768	282977	103234
50	Eagle	Colorado	39497	74456	86167	52151	17966

Top 50 cities with highest per capita income

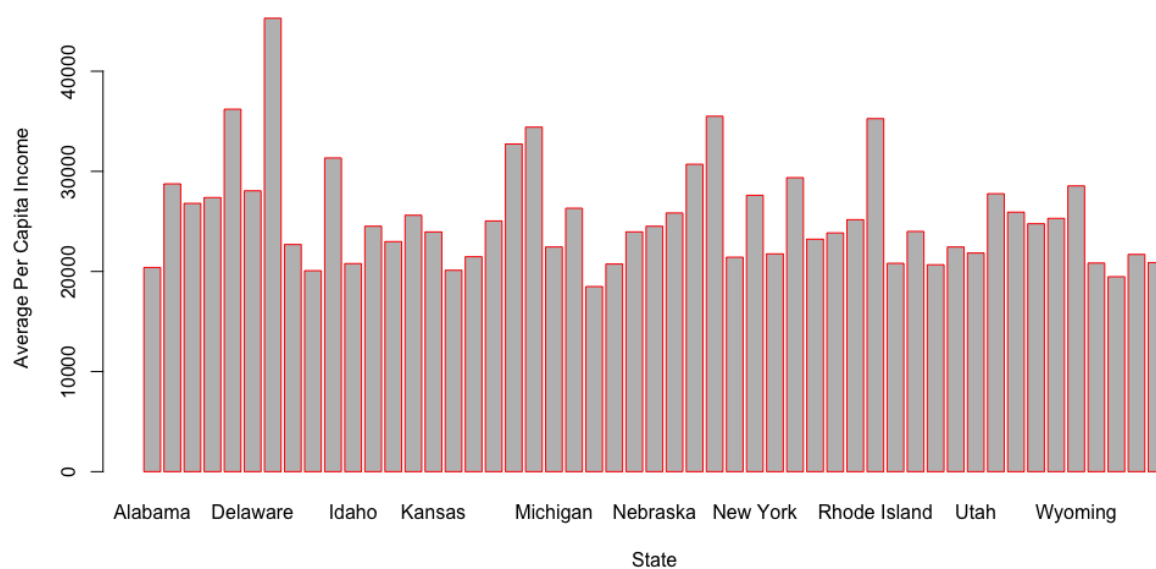
top50medianhousehold							
Rank	County-equivalent	State	Per_capita_income	Median_household_income	Median_family_income	Population	Number_of_households
19	Loudoun	Virginia	46565	122238	133550	326477	186997
3	Falls Church City	Virginia	59008	120000	152857	12731	5820
8	Fairfax County	Virginia	58532	118292	128596	1181071	389988
15	Howard	Maryland	48172	109868	128637	293821	186142
7	Los Alamos	New Mexico	51044	106686	124979	17979	7590
9	Hunterdon	New Jersey	58349	106143	125828	127047	46816
2	Arlington	Virginia	62018	103208	139244	214861	94454
27	Douglas	Colorado	43634	101591	112930	293814	183780
17	Somerset	New Jersey	47803	99020	117393	326207	115531
12	Morris	New Jersey	48814	98633	117683	495261	179665
11	Montgomery	Maryland	49038	98221	117488	989474	368563
80	Prince William	Virginia	37401	98071	106544	416668	132442
33	Nassau	New York	42400	97690	112874	1343765	441955
26	Fairfax City	Virginia	44345	97242	113538	23027	8492
92	Stafford	Virginia	36574	97110	106322	131085	42178
59	Calvert	Maryland	38260	95477	107684	89332	38923
43	Putnam	New York	48309	95117	109188	99718	34496
81	Charles	Maryland	37277	93160	104078	148957	51247
38	Santa Clara	California	41513	91702	103983	1812208	690377
4	Marin	California	56791	90839	117357	254643	102912
40	Williamson	Tennessee	41292	89779	104796	180935	66364
45	Delaware County	Ohio	48189	89757	103564	178139	64288
48	Fauquier	Virginia	39600	88409	103556	66015	23130
21	San Mateo	California	45732	88202	105844	729543	257941
86	Suffolk	New York	36945	87763	100652	1495803	497347
42	Anne Arundel	Maryland	48415	87430	101268	544426	199984
67	Sussex	New Jersey	37949	87335	100066	147924	54746
61	Queen Anne's	Maryland	38108	87256	97999	48166	17292
134	Forsyth	Georgia	34582	86569	99573	182916	58080
132	Rockwall	Texas	34617	86119	92837	81020	26820
115	Scott	Minnesota	35467	86112	97693	132774	45630
35	Chester	Pennsylvania	42210	86050	105278	583075	184788
5	Alexandria City	Virginia	54688	85706	107511	143684	65369
100	St. Mary's	Maryland	36017	85672	95186	107079	37589
18	Nantucket	Massachusetts	47331	85478	92500	10224	4869
145	Fort Bend	Texas	34084	84951	94931	68039	191486
133	Rockland	New York	34591	84951	99649	315609	98326
25	Norfolk	Massachusetts	44692	84916	108943	677296	257451
103	Carroll	Maryland	36153	84790	108643	167261	59089
87	Frederick	Maryland	36917	84570	98259	236668	86337
31	Monmouth	New Jersey	42749	84526	105971	629735	233909
29	Bergen	New Jersey	43347	83794	102222	912795	335422
78	Carver	Minnesota	37457	83773	96913	92778	33486
164	Oldham	Kentucky	33591	83391	91587	60940	19532
32	Summit	Utah	42462	83336	95879	37232	13627
69	Collin	Texas	37839	82762	97421	811308	289752
49	Hamilton	Indiana	39521	82468	99768	282977	103234
13	Fairfield	Connecticut	48721	82283	104250	926233	332655
106	Elbert	Colorado	36077	82118	98653	23295	8240
30	Middlesex	Massachusetts	42861	82090	104032	1522533	581120

Top 50 cities with highest median household income

top50medianfamily							
Rank	County-equivalent	State	Per_capita_income	Median_household_income	Median_family_income	Population	Number_of_households
3	Falls Church City	Virginia	59008	120000	152857	12731	5820
2	Arlington	Virginia	62018	103208	139244	214861	94454
19	Loudoun	Virginia	46565	122238	133550	326477	186997
15	Howard	Maryland	48172	109868	128637	293821	186142
8	Fairfax County	Virginia	58532	118292	128596	1181071	389988
9	Hunterdon	New Jersey	58349	106143	125828	127047	46816
7	Los Alamos	New Mexico	51044	106686	124979	17979	7590
12	Morris	New Jersey	48814	98633	117683	495261	179665
11	Montgomery	Maryland	49038	98221	117488	989474	368563
17	Somerset	New Jersey	47803	99020	117393	326207	115531
4	Marin	California	56791	90839	117357	254643	102912
26	Fairfax City	Virginia	44345	97242	113538	23027	8492
27	Douglas	Colorado	43634	101591	112930	293814	183780
33	Nassau	New York	42400	97690	112874	1343765	441955
43	Putnam	New York	48309	95477	107684	89332	34496
25	Norfolk	Massachusetts	44692	84916	108943	677296	257451
59	Calvert	Maryland	38260	95477	107684	89332	38923
5	Alexandria City	Virginia	54688	85706	107511	143684	65369
80	Prince William	Virginia	37401	98071	106544	416668	132442
92	Stafford	Virginia	36574	97110	106322	131085	42178
21	San Mateo	California	45732	88202	105844	729543	257941
16	Westchester	New York	47984	81946	105341	956283	343561
35	Chester	Pennsylvania	42210	86050	105278	583075	184788
40	Williamson	Tennessee	41292	89779	104796	180935	66364
13	Fairfield	Connecticut	48721	82283	104250	926233	332655
81	Charles	Maryland	37277	93160	104078	148957	51247
30	Middlesex	Massachusetts	42861	82090	104032	1522533	581120
38	Santa Clara	California	41513	91702	103983	1812208	690377
45	Delaware County	Ohio	48189	89757	103564	178139	64288
48	Fauquier	Virginia	39600	88409	103556	66015	23130
29	Bergen	New Jersey	43347	83794	102222	912795	335422
42	Anne Arundel	Maryland	48415	87430	101268	544426	199984
86	Suffolk	New York	36945	87763	100652	1495803	497347
103	Carroll	Maryland	36153	84790	108643	167261	59089
67	Sussex	New Jersey	37949	87335	100066	147924	54746
49	Hamilton	Indiana	39521	82468	99768	282977	103234
133	Rockland	New York	34591	84951	99649	315609	98326
39	Montgomery	Pennsylvania	41472	70183	99007	804621	397488
134	Forsyth	Georgia	34582	86569	99573	182916	58080
118	Denali	Alaska	35295	72500	99091	2020	716
57	Poquoson City	Virginia	38444	81892	98840	12117	4591
46	Middlesex	Connecticut	39992	76994	98501	165690	66141
275	Glasscock	Texas	31135	69107	98409	1176	409
87	Frederick	Maryland	36917	84570	98259	236668	86337
61	Queen Anne's	Maryland	38108	87256	97999	48166	17292
115	Scott	Minnesota	35467	86112	97693	132774	45630
69	Collin	Texas	37839	82762	97421	811308	289752
78	Carver	Minnesota	37457	83773	96913	92778	33486
79	Broomfield	Colorado	37442	77990	96787	57171	22016

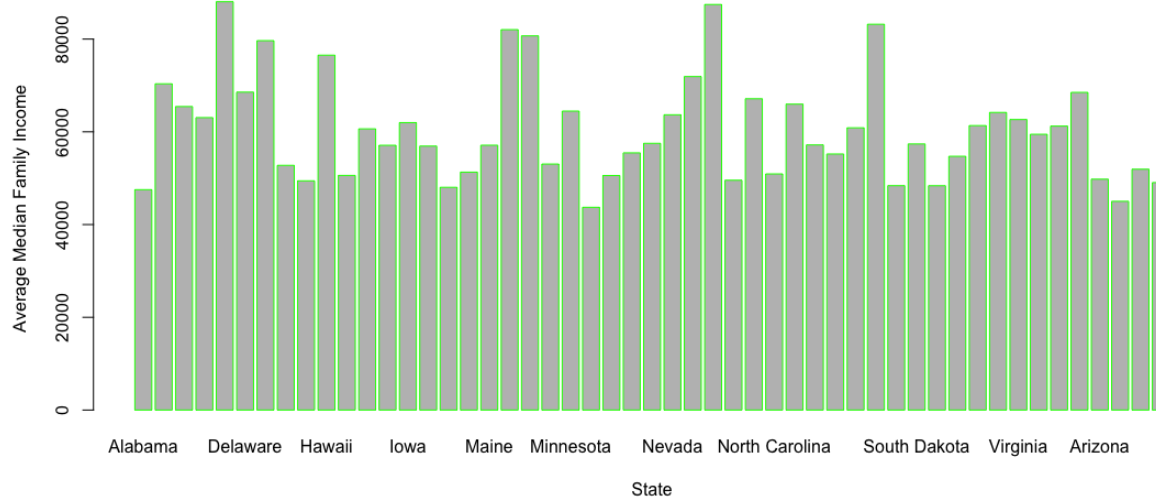
Top 50 cities with highest median family income

Average per capita income by states

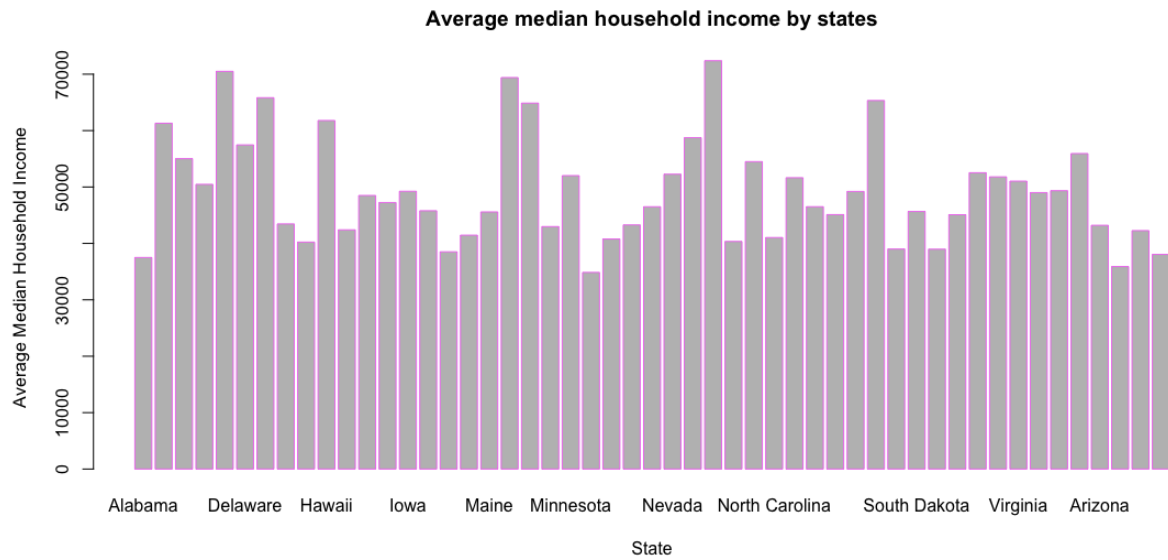


Average per capita across states

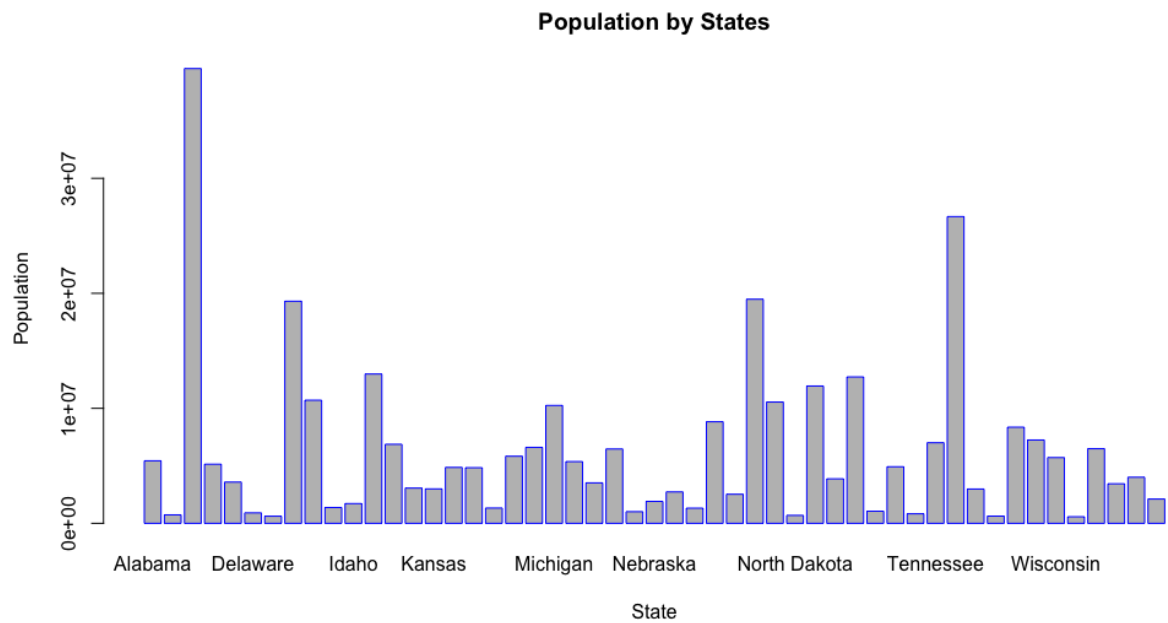
Average median family income by states



Average median family income across states



Average median household income across states



Average population across states

CONCLUSION

We understood to scrape data from websites where scraping using x-path was difficult and learnt to clean unwanted elements from the data. We learnt usage of R packages like Rcurl, XML, SQLite and performed analysis on the collected data. We obtained bar graphs and designed a SQLite database for storage.

References:

https://en.wikipedia.org/wiki/List_of_United_States_counties_by_per_capita_income

http://factfinder2.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_13_5YR_DP05&prodType=table

http://factfinder2.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_13_5YR_S1101&prodType=table