CSI 773 Final Project Report

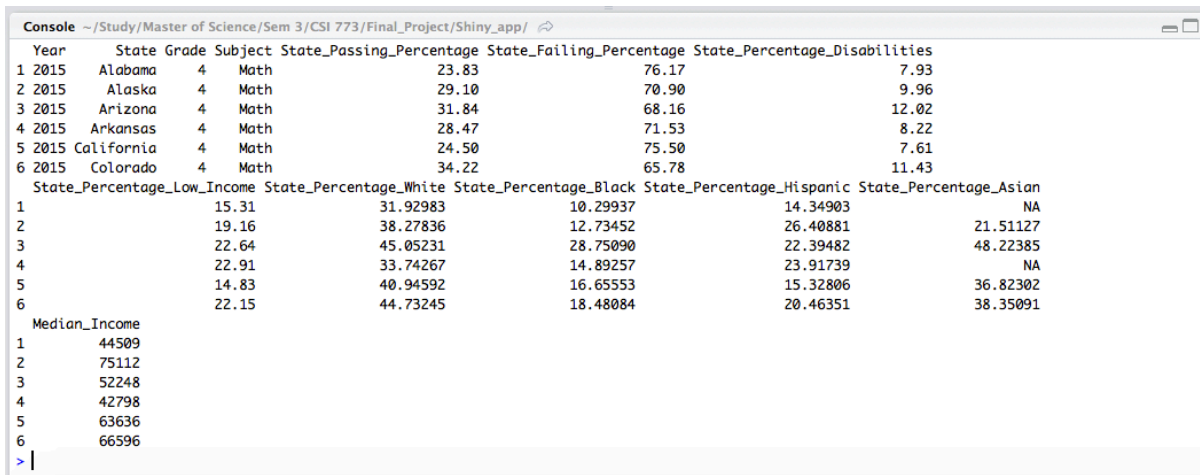# Creation Of Interactive Shiny Dashboard Using Plotly, Highcharter And Leaflet

Submitted by: Ajay Kulkarni (G01024139)

Abstract

In this project, we have built an interactive shiny dashboard using different libraries such as leaflet, plotly, and highcharter. Interactive visualizations always capture attention and help to understand details more easily. In our interactive dashboard, we have presented different visualizations like boxplots, stacked bar plot, spline chart and leaflet maps with advanced features. The dashboard helps us to give some knowledge about statewide population density, statewide median income, statewide math proficiency of 4th-grade students and top 100 schools as well as top 100 universities in the United States. This report is divided into three sections. The first section will explain about data and sources of the data. The second section is about a description of different libraries and different features which we have used for the dashboard. In the end, the third section contains different visualizations along with results.

Section 1: Data

The data which we have used for this project are imported from different sources. The purpose of our interactive dashboard is to understand the percentage proficiency in Math for 4th-grade students for different races and to study their relationship with population density and median income. The data which we have used for median income is imported from data.world website (https://data.world/garyhoov/household-income-by-state). The data for percentage proficiency in Math for 4th-grade students for all categories were downloaded from National Center for Education Statistics (https://nces.ed.gov/nationsreportcard/naepdata/) . The data for median income and Math proficiency are selected for the year 2015, and then both the datasets are combined.

```
Console ~/Study/Master of Science/Sem 3/CSI 773/Final_Project/Shiny_app/

   Year      State Grade Subject State_Passing_Percentage State_Failing_Percentage State_Percentage_Disabilities
1 2015    Alabama     4    Math                    23.83                    76.17                          7.93
2 2015     Alaska     4    Math                    29.10                    70.90                          9.96
3 2015    Arizona     4    Math                    31.84                    68.16                         12.02
4 2015   Arkansas     4    Math                    28.47                    71.53                          8.22
5 2015 California     4    Math                    24.50                    75.50                          7.61
6 2015   Colorado     4    Math                    34.22                    65.78                         11.43
  State_Percentage_Low_Income State_Percentage_White State_Percentage_Black State_Percentage_Hispanic State_Percentage_Asian
1                       15.31               31.92983               10.29937                 14.34903                      NA
2                       19.16               38.27836               12.73452                 26.40881                21.51127
3                       22.64               45.05231               28.75090                 22.39482                48.22385
4                       22.91               33.74267               14.89257                 23.91739                      NA
5                       14.83               40.94592               16.65553                 15.32806                36.82302
6                       22.15               44.73245               18.48084                 20.46351                38.35091
  Median_Income
1         44509
2         75112
3         52248
4         42798
5         63636
6         66596
> |
```

Figure 1: Percentage proficiency in Math and monthly income data

The variables in the data are as follows,
1) **Year**: Year of the data.
2) **State**: Names of the all states in the United States.
3) **Grade**: Grade of the students.

4) **Subject**: Name of the subject.
5) **State_Passing_Percentage**: Statewide average proficient percentage of the students in Math.
6) **State_Failing_Percentage**: Statewide average failing percentage of the students in Math.
7) **State_Percentage_Disabilities**: Statewide proficient percentage of the students in Math who have disabilities.
8) **State_Percentage_Low_Income**: Statewide proficient percentage of the students in Math who belongs from low income family.
9) **State_Percentage_White**: Statewide proficient percentage of the students in Math who are white.
10) **State_Percentage_Black**: Statewide proficient percentage of the students in Math who are black.
11) **State_Percentage_Hispanic**: Statewide proficient percentage of the students in Math who are Hispanic.
12) **State_Percentage_Asian**: Statewide proficient percentage of the students in Math who are Asian.
13) **Median_Income**: Statewide median income.

We also have used data for top 100 schools and top 100 universities in the United States. The source of the data for top 100 schools is neighborhoodscout.com (https://www.neighborhoodscout.com/blog/top-100-best-schools) and the source for top 100 universities is 4icu.org (http://www.4icu.org/us/). We have used this data for plotting on a leaflet. Thus, it was very important to have latitude and longitude for every school and university. For finding the exact location in terms of latitude and longitude, we have used a batch geocoding website (https://www.doogal.co.uk/BatchGeocoding.php) and then stored all the results in a csv file.

```
Console ~/Study/Master of Science/Sem 3/CSI 773/Final_Project/Shiny_app/
> head(data1)
  School_Rank              School_Name    City_State     lat       lon                                               School_Url
1        100 Meadow Glens Elementary School Naperville, IL 41.74569  -88.10718         https://www.naperville203.org/meadowglens
2         99        Bellevue High School   Bellevue, WA 47.60414 -122.19870                      http://www.bsd405.org/bhs/
3         98      Village Elementary School   Syosset, NY 40.82106  -73.49703 http://www.syossetschools.org/schools/village_elementary
4         97          Ps 188 Kingsbury    Flushing, NY 40.73610  -73.75362 http://schools.nyc.gov/SchoolPortals/26/Q188/default.htm
5         96       Lexington High School   Lexington, IL 40.64504  -88.78023                  http://www.lexington.k12.il.us/
6         95 Daves Creek Elementary School    Cumming, GA 34.13632  -84.14261        https://www.forsyth.k12.ga.us/davescreek
  University_Rank                    University_Name    ulat      ulon             University_Url
1              1 Massachusetts Institute of Technology 42.36009  -71.09416        http://web.mit.edu/
2              2                Stanford University 37.42748 -122.16972   https://www.stanford.edu/
3              3                 Harvard University 42.37700  -71.11666    https://www.harvard.edu/
4              4   University of California, Berkeley 37.87190 -122.25854     http://www.berkeley.edu/
5              5           University of Washington 47.65534 -122.30352 https://www.washington.edu/
6              6                  Cornell University 42.45345  -76.47350    https://www.cornell.edu/
> |
```
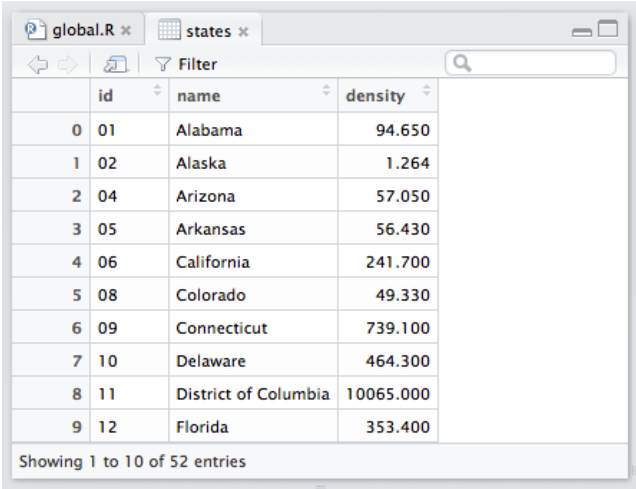
Figure 2: Top 100 schools and universities in the United States

The description about variables are given below,
1) **School_Rank**: Rank of the school.
2) **School_Name**: Name of the school.
3) **City_State**: City and state of the school.
4) **lat**: Latitude position of the school.
5) **lon**: Longitude position of the school.

6) **School_Url**: School website.
7) **University_Rank**: Rank of the university.
8) **University_Name**: Name of the university.
9) **ulat**: Latitude position of the university.
10) **ulon**: Longitude position of the university.
11) **University_Url**: University website.

The data which we have used for population density and shapes of US states on choropleth was in GeojSON format, and we imported that data from leafletjs.com. After importing GeojSON data from the website, we converted that data in R using geojson and geojsonio libraries (http://leafletjs.com/examples/choropleth/us-states.js).

Figure 3: States and population density data of United States

The description about variables are given below,
1) **id**: Identification number of every record.
2) **name**: Name of the state.
3) **density**: Population density.

So, in this way data from different source are combined together to build an interactive dashboard.

Part 2: Libraries used and Shiny dashboard

In this project, we built a shiny dashboard using plotly, highcharter and leaflet. The purpose of using different libraries was to learn and implement them. The description of three libraries and the shiny dashboard is given on next page.

1) Plotly

Plotly is a graphing library in R which is a free and open source. Plotly can be used for creating interactive, publication-quality graphs online. Plotly supports different plots such as line plots, scatter plots, area charts, histograms, etc., For this project, we have used plotly for plotting scatterplot of median income vs. percentage proficiency in math and box plot for every category.

2) Highcharter

Highcharts is a popular JavaScript graphing library which can be used in R using Highcharter. Higcharter provides a rich R interface and brings all the highcharts capabilities. It is used to produce interactive web visualizations. There are various chart types like scatters, bubble, line, time series, heat maps, networks, etc., which are available. For this project, we have used highcharter for stacked column chart and spline chart. We have represented statewide proficiency and below proficiency percentage in Math using stacked column chart. A spline chart is used to denote the statewide median income. We have used two y-axes in one plot, where one of the y-axes represent the percentage while another y-axis represents the median income in dollars.

3) Leaflet

The leaflet is an open-source JavaScript library for creating interactive maps. In this project, we have created maps with choropleths. We are providing three options for selecting choropleths, and these options are population density, percentage proficiency in Math & Median income. We also have used customized markers to represent schools and universities. In addition to that, we are also providing customized popups for every marker containing URL of the respective school or university. To get more analytical clarity, we have clustered the markers which will help us to understand more about the distribution of top 100 schools and universities all over the United States. Leaflet also provides a feature called Layer Groups and Layers Control. We have used this feature on our map, and we are giving the user a choice to select or deselect schools and universities. In addition to that we also have used a grey basemap to focus only on United States and not on any other country.

4) Shiny dashboard

The shiny dashboard can be built in R using "shinydashboard" package in R. Shiny dashboard provide different themes which makes it easy to create an interactive dashboard. The structure of shiny dashboard consists of header, sidebar, and body. The simple dashboard can be easily built using a single file, but for this project, we have used three files for creation. The three files which we used are global.R, ui.R and server.R. The file global.R consist of all the data imports, ui.R contain details about the user interface, positions of the plots as well as other features which we want to see on the dashboard. The server.R file contains the code for creating different plots. All the files are connected and are necessary for deploying the shiny

dashboard. The theme which we have used for the shiny dashboard is "black" and we also have provided a link in the header of the dashboard. One of the most important advantages of the shiny is, it provides free web hosting of a shiny app for certain period. So, this project is also hosted on shiny server to make complete web interactive

Part 3: Results and Visualizations

In this section, we will present the visualizations and analyses of the results. We will start with box plots and scatter plot. After that, we will analyze stacked column chart overlapped by the spline chart, and in the end, we will examine the leaflet maps.

1) Boxplots and scatter plot
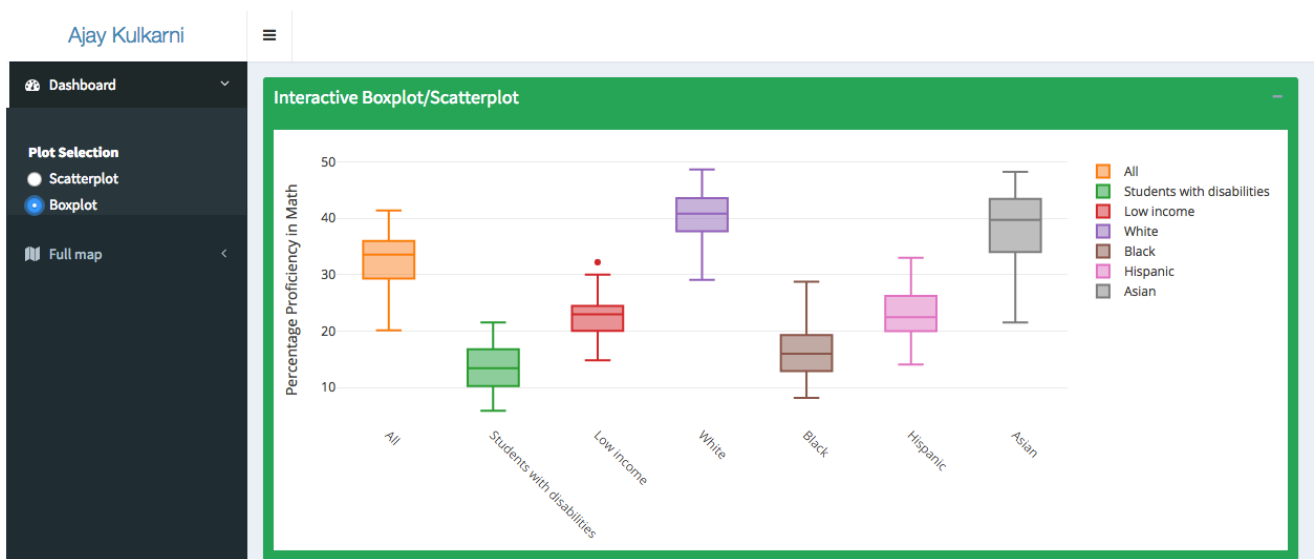
Figure 4: Boxplots for every category with sidebar

We have provided control on the shiny dashboard to select box plots or scatterplot. So a user can easily choose what they want to see. Figure 4 represents box plot for every category. Boxplots are used to easily understand about a min, max, median and other information such as quartiles. In the above figure "all" indicates the average percentage proficiency in Math for all the students. Average minimum proficiency in Math for 4th-grade students is around 20%, and Average maximum proficiency in Math for 4th-grade students is approximately 41%. Also, it can be observed from the boxplots that 4th-grade Students with disabilities have the minimum proficiency in Math as compared to other categories. The maximum percentage proficiency is seen for White and Asian categories.

Figure 5 shows the scatterplots which we have plotted to understand median income and proficiency in Math. It can be observed that Puerto Rico shows the minimum, median income, and percentage proficiency in Math. On the other hand, New Hampshire shows the maximum

median income and percentage proficiency in Math. It also can be observed that majority of the states have median income 40-60K and percentage proficiency in Math is 25%-35%.
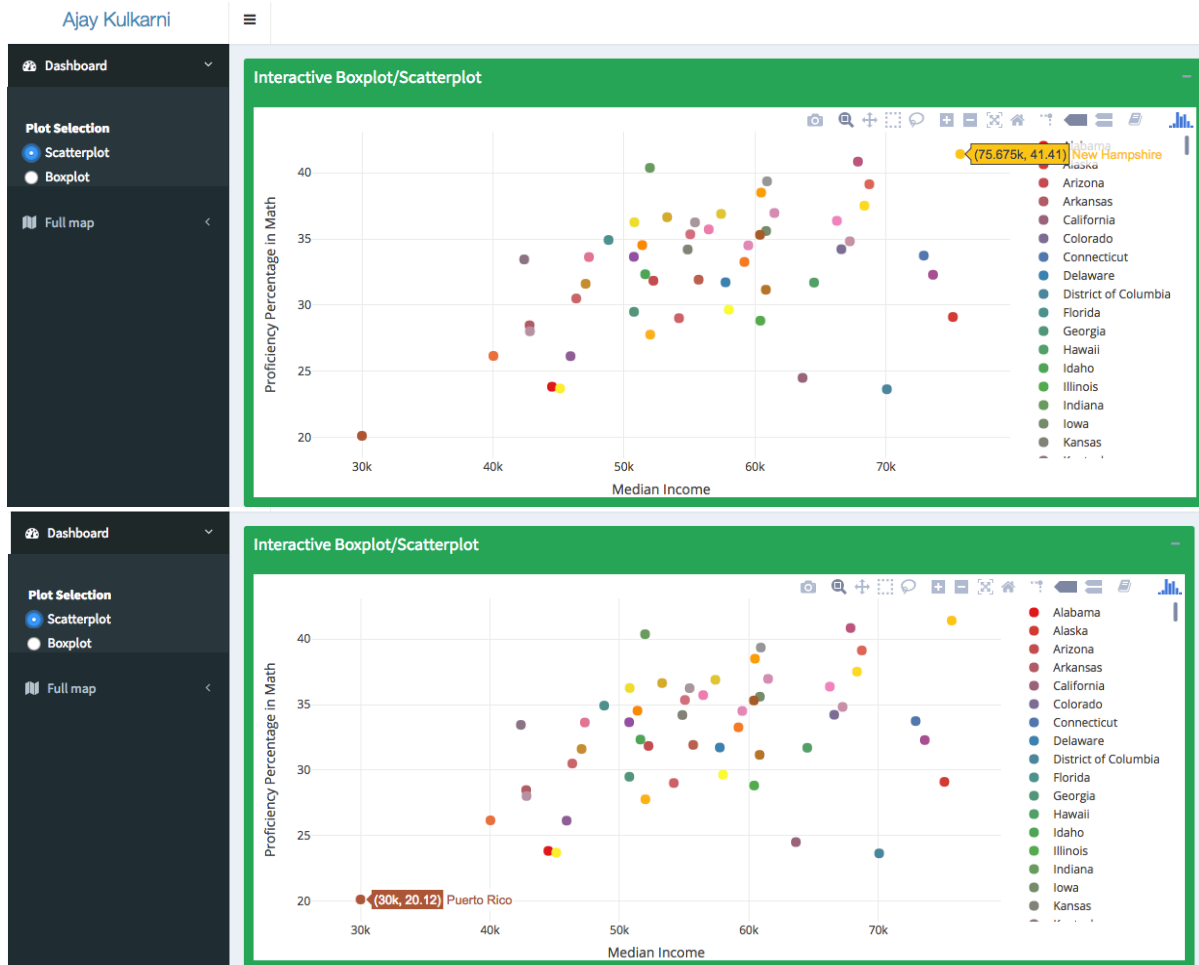

Figure 5: Scatterplots indicating minimum and maximum Proficiency Percentage in Math and Median Income

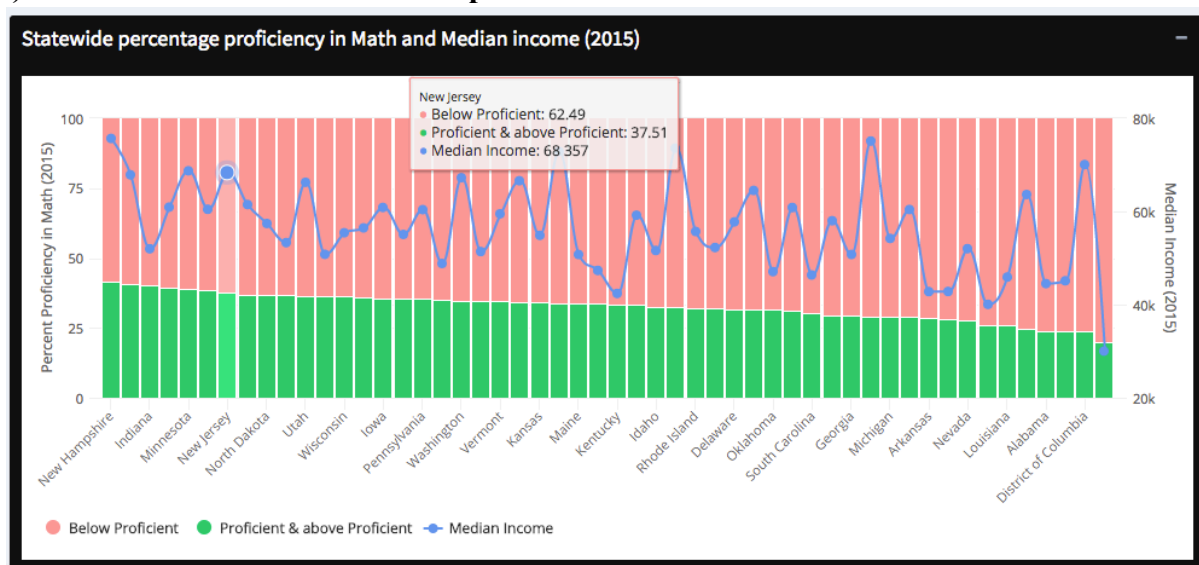## 2) Stacked column chart and line plo

Figure 6: Stacked Colum chart and line plot

Figure 6 represents the stacked column chart and spline chart. In figure 6 red color indicates the below proficient percentage of the 4th-grade students for every state and green color indicates the above proficient percentage of the 4th-grade students. Also, the spline chart shows the median income for every state. From the spline chart, it can be seen that highest median income is observed for New Hampshire, but Alaska also have approximately same median income. In case of proficiency in Math New Hampshire have around 11% more proficiency as compared to Alaska.

3) Leaflet maps with choropleth

We have used leaflet for analyzing statewide population density, median income and percent proficiency in Math. We have used choropleths for representing the above three variables. We also have used customized markers to represent top hundred schools and universities. For clear visualization and understanding, we also have provided a legend for all the variables.
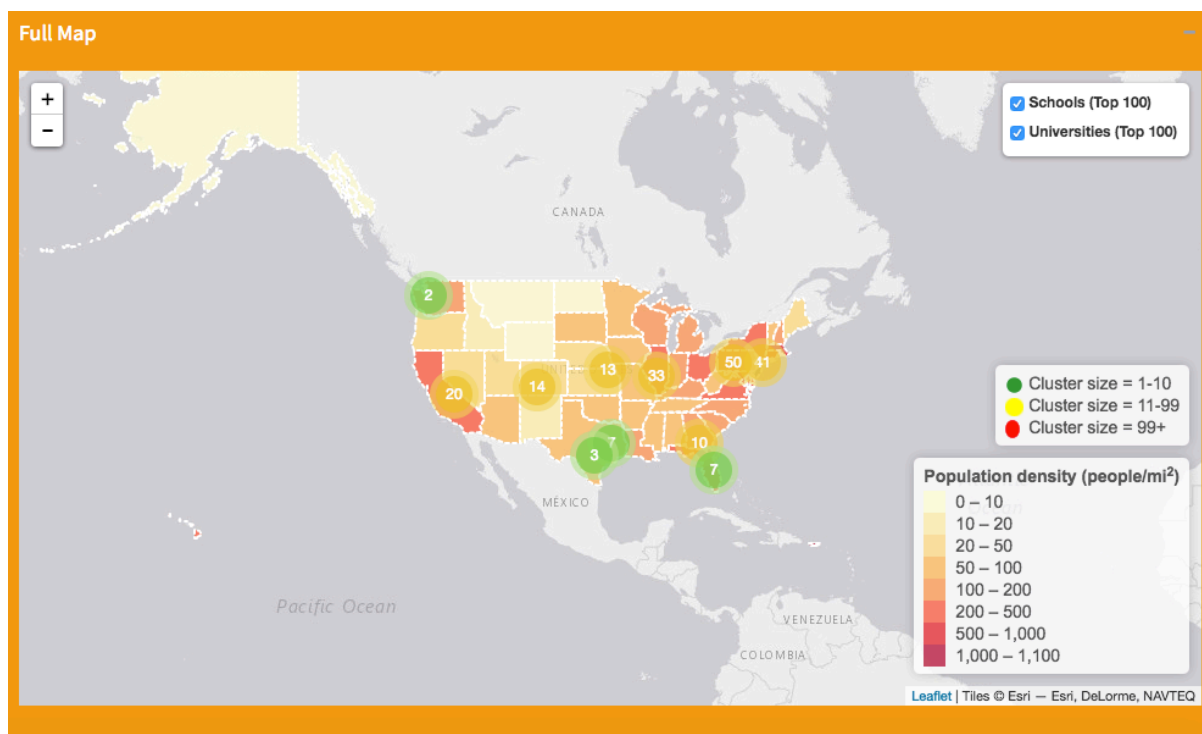

Figure 7: Leaflet for Population Density

Leaflet with choropleth is shown for the population density and shading of orange color indicates the population density. It shows that darker the shade more the population density and vice versa. Also, the above leaflet is displaying a menu at the top for selecting schools or universities. So if users want to see only schools, then they can deselect the university checkbox. We also have clustered the schools and universities to understand the which part of United States have more top schools and universities. So it is observed that east coast has more number of top schools and universities as compared to central as well as the west coast. Also,

it can be observed that population density in the central United States is less as compared to east and west part of United States.
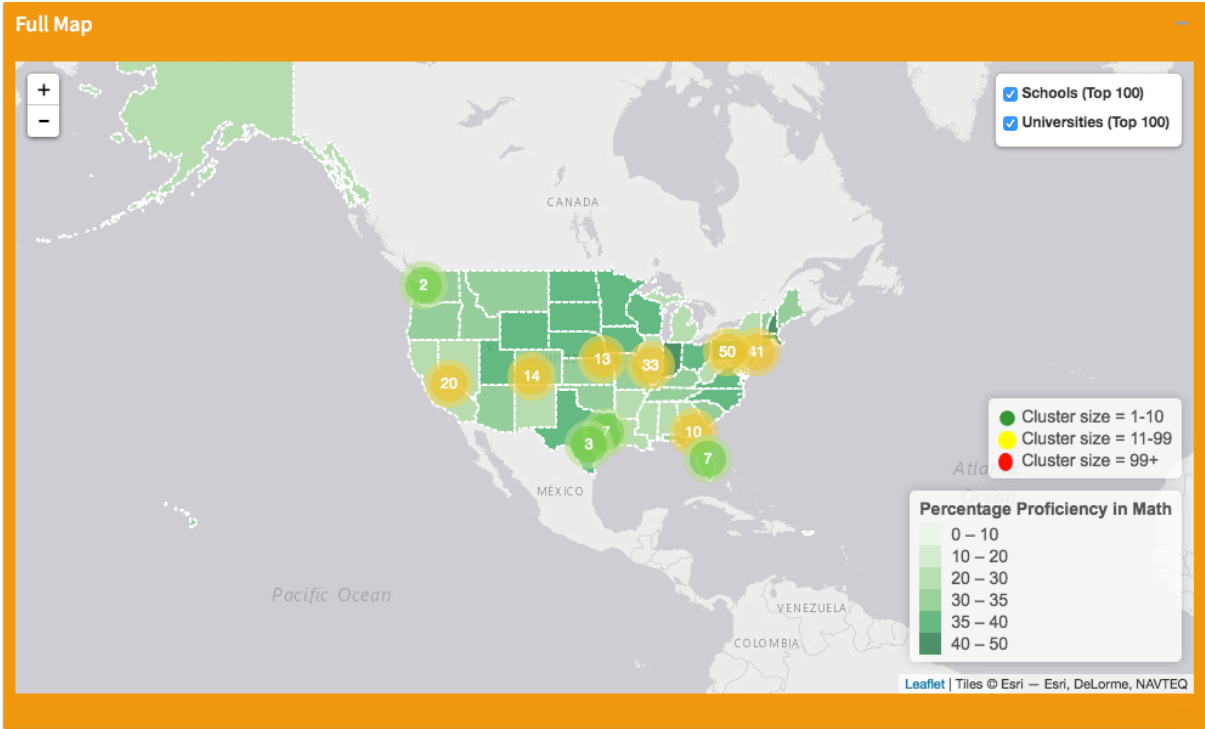

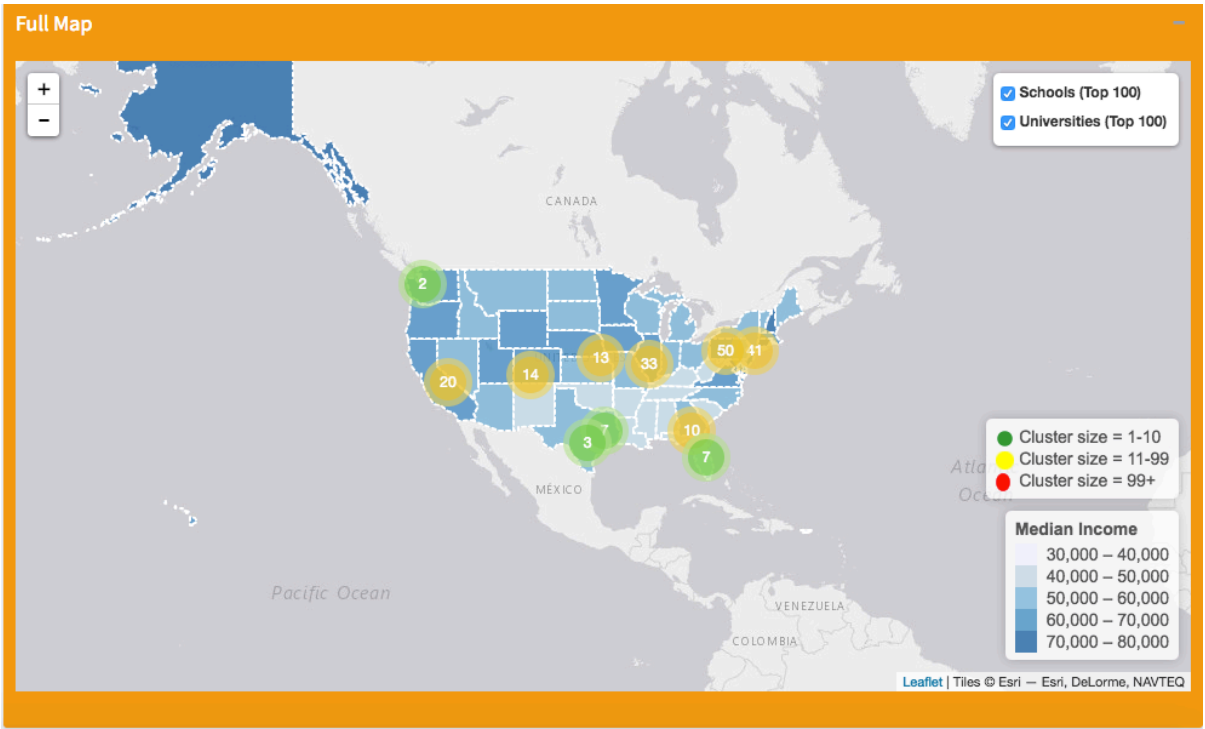Figure 8: Leaflet for Percentage Proficiency in Math

Figure 9: Leaflet for Median Income

The other two leaflet maps with different choropleths for Percentage Proficiency in Math and Median Income are shown above. We also have we provided tooltip for every state. So, whenever a user moves cursor over the state, they will get the information about all the

variables. In addition to that, we also have provided a popup window for every customized marker. When we click the marker, we will get a link to that respective school or university in the popup as shown in the figure 8. From figure 7 it can be observed that Percentage Proficiency in Math of 4th-grade student is more in the central part of United States as compared to the east and west part of the United States. Also, it can also be observed that Median Income of the majority of states of the central United States lies within 40-60K.

Conclusion

In this project, we have successfully built an interactive shiny dashboard and published on a shiny server. This dashboard was built using plotly, higcharter and leaflet libraries in R. From this dashboard we understand the statewide distribution of Population Density, Percentage Proficiency in Math of 4th-grade students and Median Income. This dashboard also helped us to understand the distribution of top hundred schools and universities in the United States.

References

Data Sources

1) https://data.world/garyhoov/household-income-by-state
2) https://nces.ed.gov/nationsreportcard/naepdata/
3) https://www.neighborhoodscout.com/blog/top-100-best-schools
4) http://www.4icu.org/us/

Other references

1) https://rstudio.github.io/leaflet/
2) https://plot.ly/r/
3) http://jkunst.com/highcharter/
4) https://rstudio.github.io/shinydashboard/
5) https://shiny.rstudio.com/
6) https://www.doogal.co.uk/BatchGeocoding.php
7) https://stackoverflow.com/questions/31440564/adding-a-company-logo-to-shinydashboard-header
8) https://shiny.rstudio.com/articles/shinyapps.html
9) https://stackoverflow.com/questions/40084416/two-y-axis-in-highcharter-in-r
10) https://upload.wikimedia.org/wikipedia/commons/0/0e/Ski_trail_rating_symbol-green_circle.svg
11) http://www.clker.com/cliparts/o/b/y/x/Z/c/yellow-dot-md.png

12) http://www.clker.com/cliparts/X/9/P/m/2/g/transparent-red-circle.svg

13) http://www.r-chart.com/2010/05/number-formatting.html

Note: I have changed the line plot and replaced it with spline chart. I also have changed the color of the spline but was not able to change the color of the second y-axis and labels. To change those colors some changes were needed to make using JavaScript. In case of leaflet I searched about different Projections but due to technical complexity it was not feasible to implement that. For clustering, leaflet uses a technique of grouping and groups the markers that are close to each other on each zoom level. The shiny app is uploaded on shiny server and link for the app is https://ajayk.shinyapps.io/csi_773/.