Project Report

on

Diversity at Clemson University

Part I - Data Analysed

For the implementation of this project I collected the data of current graduate students at Clemson University. This data was provided to me by Miranda Hix, the Director of Web Technologies at Clemson University. Originally this data had the nation of the students (arranged in the alphabetical order), their program of study and no. of students from each category. But for the purpose of this implementation I reduced the data to the nation of the students and the total no. of students who belonged to the category and in addition to that I manually added the latitude and longitude of each country in the data set. Below is a snap shot of how the data looks like,

|	CITZ_NATION	STUDENT_COUNT	LATITUDE	LONGITUDE
2	Afghanistan	2	34.5	69.1
3	Chile	2	-33.4	-70.6
4	China	605	35	103
5	Colombia	11	4.5	-74.06
6	Costa Rica	10	9.9	-84
7	Croatia	1	45.8	16
8	Cyprus	1	35	33
9	Denmark	1	55.72	12.57
10	Dominican Republic	2	19	-70.6
11	Ecuador	3	-0.15	-78.35
12	Egypt	2	26	30
13	El Salvador	8	13.6	-89.1
14	Ethiopia	1	9	38.7
15	France	3	47	2
16	Germany	12	52.5	13.3
17	Ghana	4	5.5	-0.2
18	Greece	3	39	22
19	Haiti	1	18.5	-72.3
20	Honduras	2	14.1	-87.2
21	Hong Kong	1	22.2	114.1
22	India	616	21	78
23	Iraq	15	33.3	44.4
24	Israel	2	31	35
25	Italy	2	41.9	12.4
26	Jamaica	4	18.1	-77.32
27	Japan	4	35.6	139.7
28	Jordan	13	31.9	35.9
29	Kazakhstan	1	48	68
30	Kenya	8	-1.2	36.8
31	Korea Republic of	21	37 5	126 9

Part II - Tools Used

I used Processing for the implementation firstly because I was familiar with it and secondly because it had this amazing library called Unfolding Maps. It is a library that enables you to create interactive maps. It comes with a couple of existing map providers. I used *Google.GoogleMapProvider();* for my project as it displayed all the countries which is exactly what I need for the implementation.

Part III - Describe what characteristics of the data you learned about.

1) Clemson on the world map - I realized that the graduate students at Clemson are from almost every part of the world.

2) Maximum no. of students - The maximum no. of students are from US (of course!)
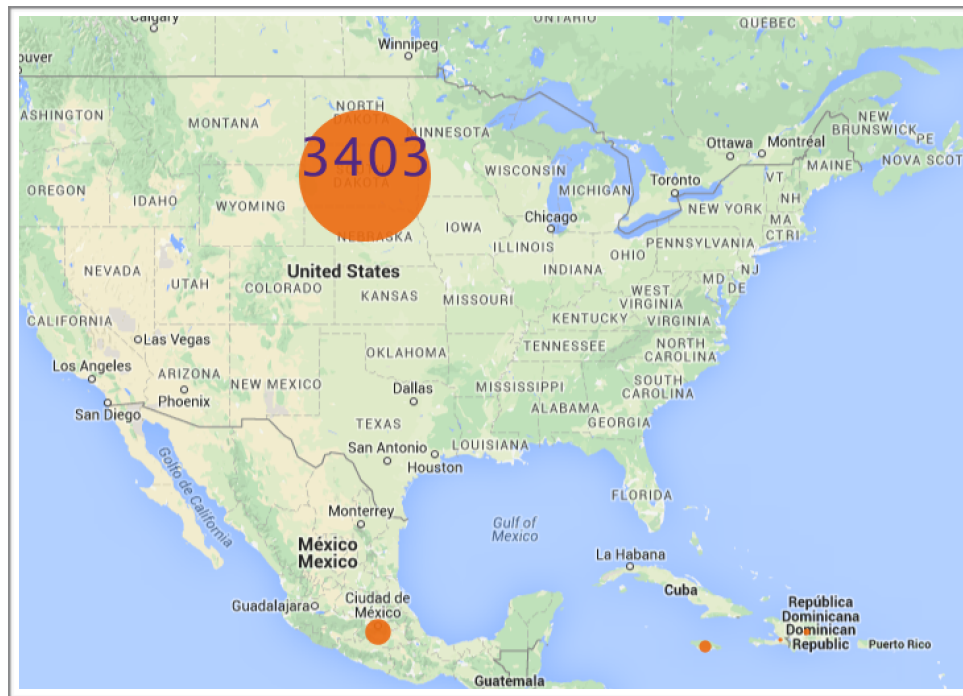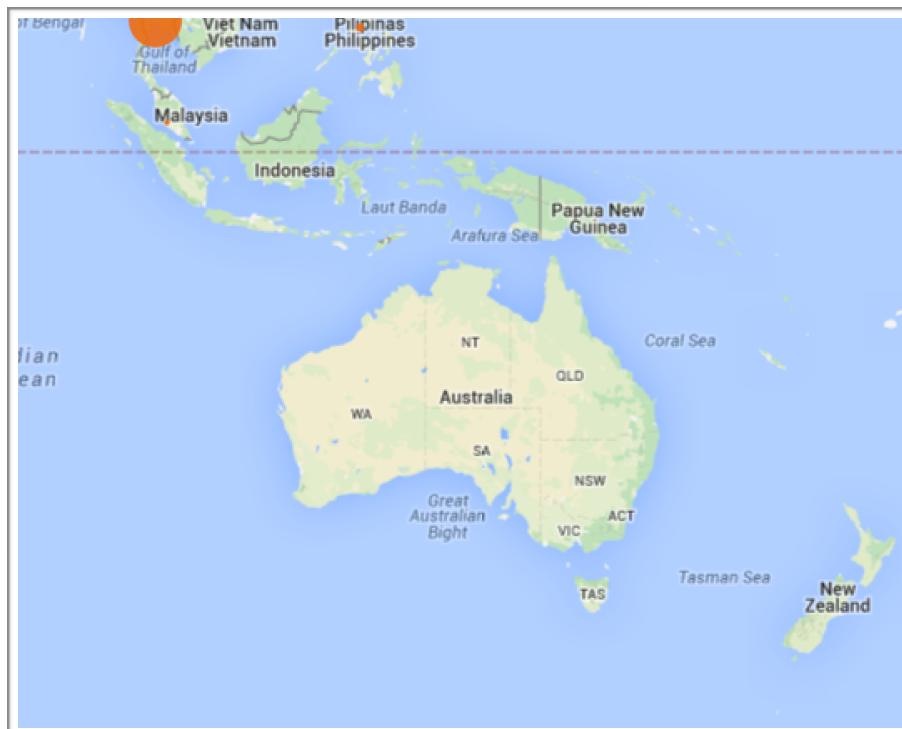

3) Maximum number of the international students are either from India or China.

4) Surprisingly, there is not even a single student from Australia or New Zealand.

Part IV - Lessons learned while implementation

People like looking at visual representations rather than a plain boring .csv file (high five to all the people working with data visualisation). Scrolling through the csv or excel file to search for outliers is almost impossible. May be finding out that the maximum number of students are from US would not have been so difficult but discovering that there are no students from Australia would have been really difficult or could be missed easily.

Part V- Challenges faced while implementation

The most challenging and time consuming part of this project for me was to get this kind of a data. It would have been difficult to plot a map without using Unfolding maps. The latitudes and longitudes had to be used with their signs. At first I did not take the sign into account and ended up with a map that was not plotted properly. But once I figured that out the later part was pretty simple.

Part VI- Evaluation of the final visualization

A map seems like the best technique to represent a data like this because anything else would not have done justice to this type of a data. I used circular markers to plot a point on the map. They had to be orange in color because it's Clemson we are talking about. For the points I used the color orange and for the no. of. students at each location I used Regalia (Goes perfectly well with Orange). If you click on a point on the map then the number of students at that location is displayed. You can zoom in and out of the map using the scroll or simply by clicking at the location and move around the map by sliding through it.