Analysis of Crime Statistics of U.S.A

OVER THE YEARS 2010-2013

CHAITANYA KRISHNA SAI KOSARAJU | STAT 515 | May 6, 2016 | G00989798

Professor Dr. Daniel B. Carr

2. Introduction:

The objective of this project is to show a dataset in different map forms and plots while comparing the data available in it. The dataset is about different types of Violent Crime Rates of 52 states in the U.S.A from the year 2010 to 2013. The source of the data set is www.fbi.gov. The dataset was downloaded in .csv format. For the purpose of comparing the data. I have normalized the values by dividing the total crimes per attribute of a state by its population. The original dataset looked like below:

3. Data Set:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Order	State	vcrime_A	murder_A	Rape_A	Robbery_	Assault_A	vcrime_B	murder_B	Rape_B	Robbery_I	Assault_B	vcrime_C	murder_C	Rape_C
2	1	Alabama	377.7614	5.711613	28.22332	99.60801	244.2185	422.0735	6.297419	28.68359	102.7253	284.3672	453.8535	7.155207	27.11447
3	2	Alaska	638.8062	4.364777	75.04601	83.63476	475.7607	617.1232	4.083179	59.13569	81.10037	472.8039	621.2063	4.223978	82.08597
4	3	Arizona	408.0872	6.398606	33.87037	108.526	259.2922	411.6228	6.336028	35.41918	111.4985	258.3692	439.736	5.600736	35.62256
5	4	Arkansas	505.3297	4.732643	44.99441	81.34659	374.2561	484.5472	5.555712	41.59925	83.26709	354.1252	474.4646	5.932951	42.76526
6	5	California	440.5787	4.85586	22.36273	155.9995	257.3606	415.9129	4.810227	20.56963	145.7349	244.7982	432.0185	5.057181	21.03669
7	6	Colorado	320.7869	2.386067	43.7048	62.27636	212.4196	325.7578	2.9627	45.29551	65.73615	211.7635	318.5996	3.221191	42.01467
8	7	Connectic	281.3858	3.637282	16.31181	99.40973	162.0269	273.2718	3.581324	19.19366	102.8791	147.6177	284.2676	4.084948	25.71279
9	8	Delaware	620.8697	5.345604	34.74643	203.6898	377.0878	565.1863	4.566037	32.18499	171.2821	357.1532	559.0611	6.347905	27.06212
10	9	District of	1330.18	21.937	31.07742	718.7693	558.3965	1234.621	17.94846	28.75077	680.2133	507.7087	1306.914	14.62467	39.2207
11	10	Florida	542.3505	5.249634	28.5778	138.7457	369.7774	522.2987	5.233678	28.05124	136.2777	352.7361	500.4279	5.366647	27.97677
12	11	Georgia	403.3175	5.759909	21.60482	127.7193	248.2335	378.1514	5.718619	21.19192	125.3967	225.8442	388.03	5.997325	21.92482
13	12	Hawaii	262.736	1.764315	26.8323	77.48285	156.6565	290.3034	1.249723	31.9047	76.60069	180.5483	244.7988	2.131881	20.95125
14	13	Idaho	221.0411	1.339643	33.49107	13.6516	172.5588	203.1154	2.296531	27.74974	11.73782	161.3313	211.6636	1.849983	30.55662
15	14	Illinois	435.1695	5.502457	23.63874	156.2978	249.7305	430.5867	5.619365	28.89959	157.8566	238.2112	416.2149	5.798623	27.82404
16	15	Indiana	314.4606	4.50353	27.15999	95.91595	186.8811	333.5389	4.811991	27.0983	107.622	194.0065	348.5918	4.781145	25.71022
17	16	Iowa	273.54	1.280218	27.44263	33.1872	211.63	256.8972	1.510001	27.37698	27.08154	200.9287	266.2854	1.477175	28.59155
18	17	Kansas	369.105	3.504937	38.79966	54.11623	272.6841	356.1717	3.855431	38.02857	51.13704	263.1507	358.6252	2.944147	36.90699
19	18	Kentucky	242.616	4.309384	31.82492	86.37204	120.1097	239.8046	3.52586	33.71459	85.10458	117.4595	224.7332	4.493743	29.31303

FinalDAtaStat

I have also taken the number of Law Enforcement Employees per State for the year 2010 for further analyzing the relationship between number of police personnel present and the crime rates, and find if there exists a trend between these entities.

|	State	Total_Law_Enforcement_Employees_2011	vcrime_2011
2	ALABAMA	1549	377.7614496
3	ALASKA	623	638.8062475
4	ARIZONA	1837	408.08715
5	ARKANSAS	34	505.3297109
5	CALIFORNIA	11681	440.5787133
7	COLORADO	1288	320.7868614
8	CONNECTICUT	1561	281.3857598
9	DELAWARE	1559	620.8696853
0	FLORIDA	5148	1330.180166
1	GEORGIA	2555	542.3505064
2	IDAHO	455	403.3175012
3	ILLINOIS	3550	262.735968
4	INDIANA	1722	221.0410683
5	IOWA	927	435.1695224
6	KANSAS	1280	314.4605588
7	KENTUCKY	1894	273.5400175
8	LOUISIANA	1663	369.1049582
9	MAINE	476	242.6160313
0	MARYLAND	3857	548.9512001
1	MASSACHUSETTS	2877	122.0300807
2	MICHIGAN	2309	547.669788
3	MINNESOTA	813	466.6269271
4	MISSISSIPPI	76	490.3051912
5	MISSOURI	2567	235.9573335
6	MONTANA	324	269.7067398
7	NEBRASKA	702	455.0207750

Full-time_state_law_enforcement

I have faced some errors while executing the project. Initially, R wasn't accepting numerical values for column names when executing Micromaps. I also had to change the abbreviations of state names. The major problem I found was to understand that the data set was supposed to be normalized before analyzing, as it is obvious that states with huge population can have higher rates, when compared to low populated states. And, it is meaningless to compare and analyze this data, as always highly populated states like California and Utah, would be topping the list, no matter there are states with higher crime rates per population.

4. Histogram:

4.1 Histogram for violent crimes in 2010

From the Histogram, it is evident that D.C has topped the crimes for 2010, with an astonishing crime rate of 1330. Second highest rate is in Alaska rounding up to 640.
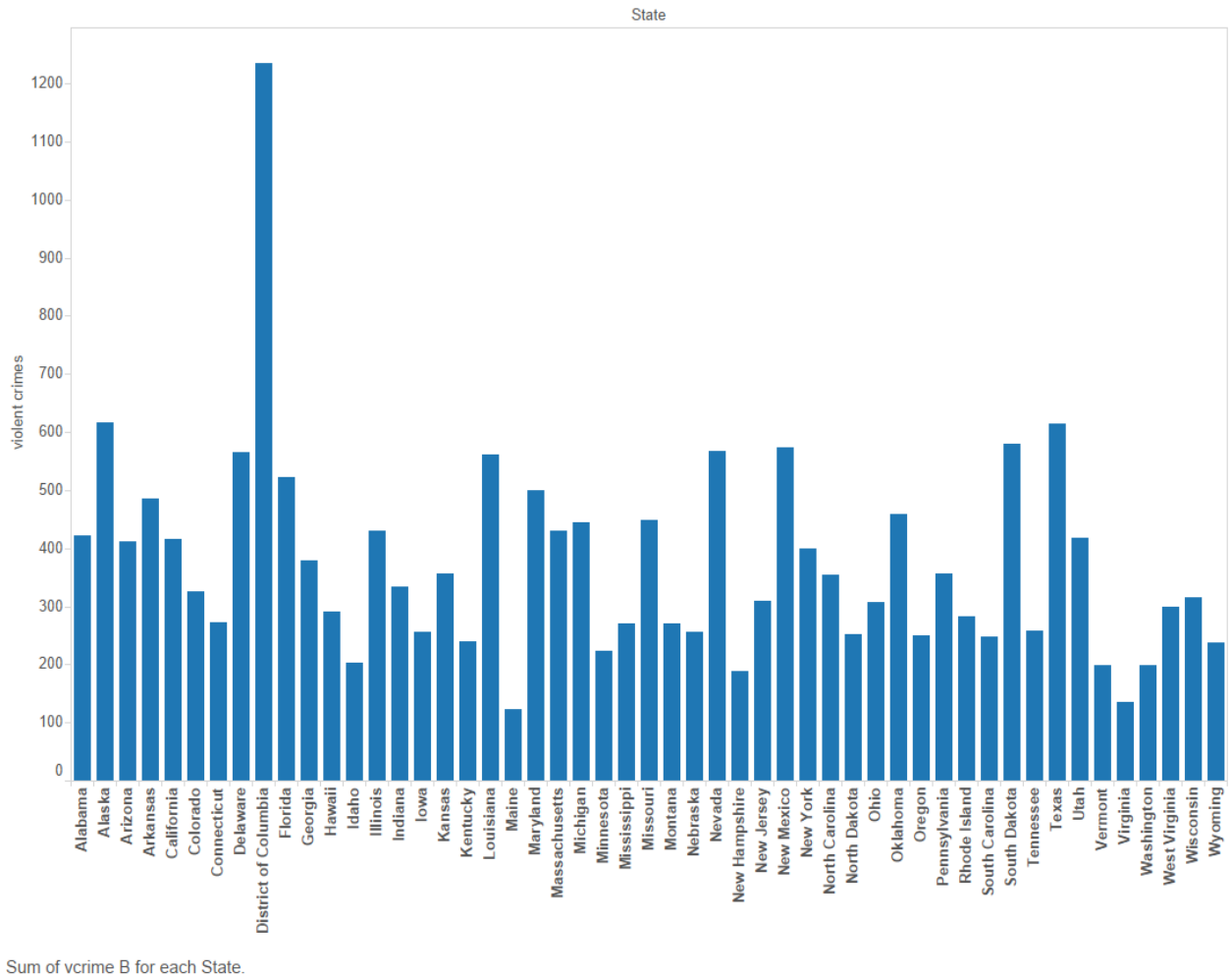
Sheet 1
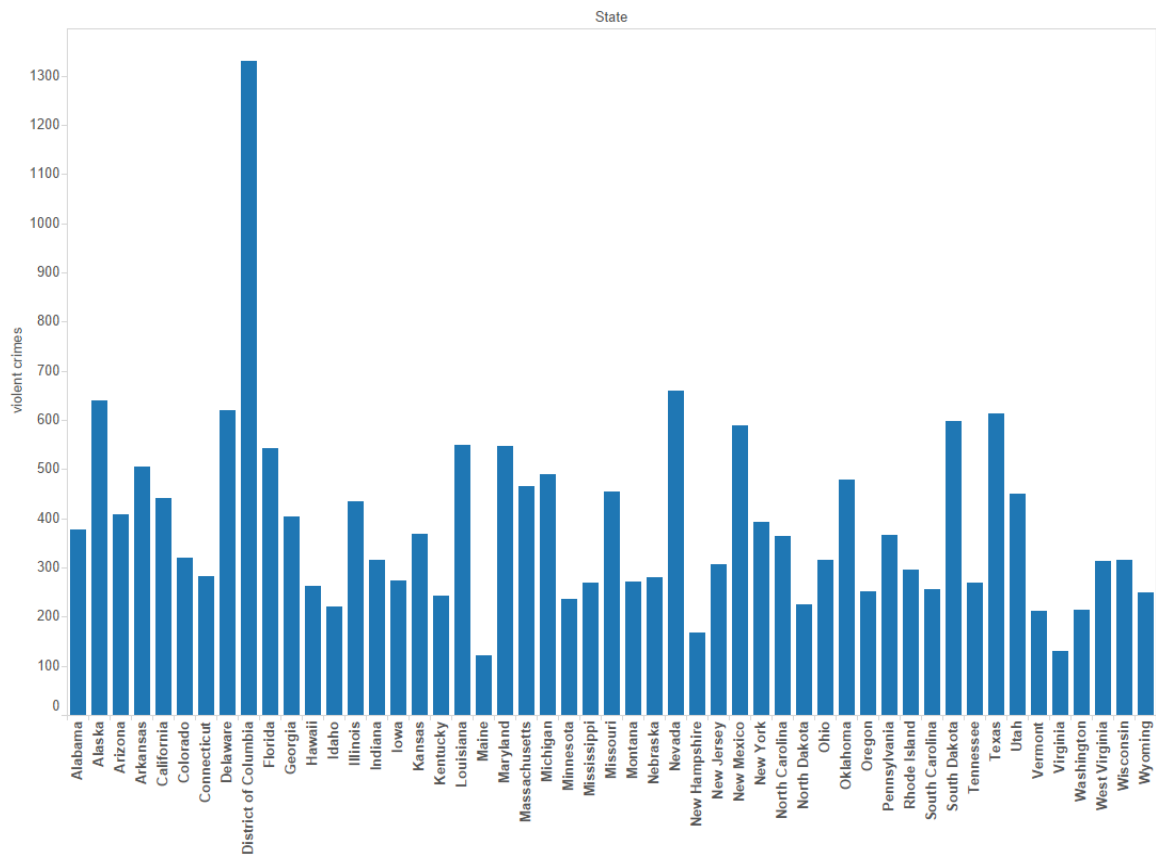
Sum of vcrime A for each State.
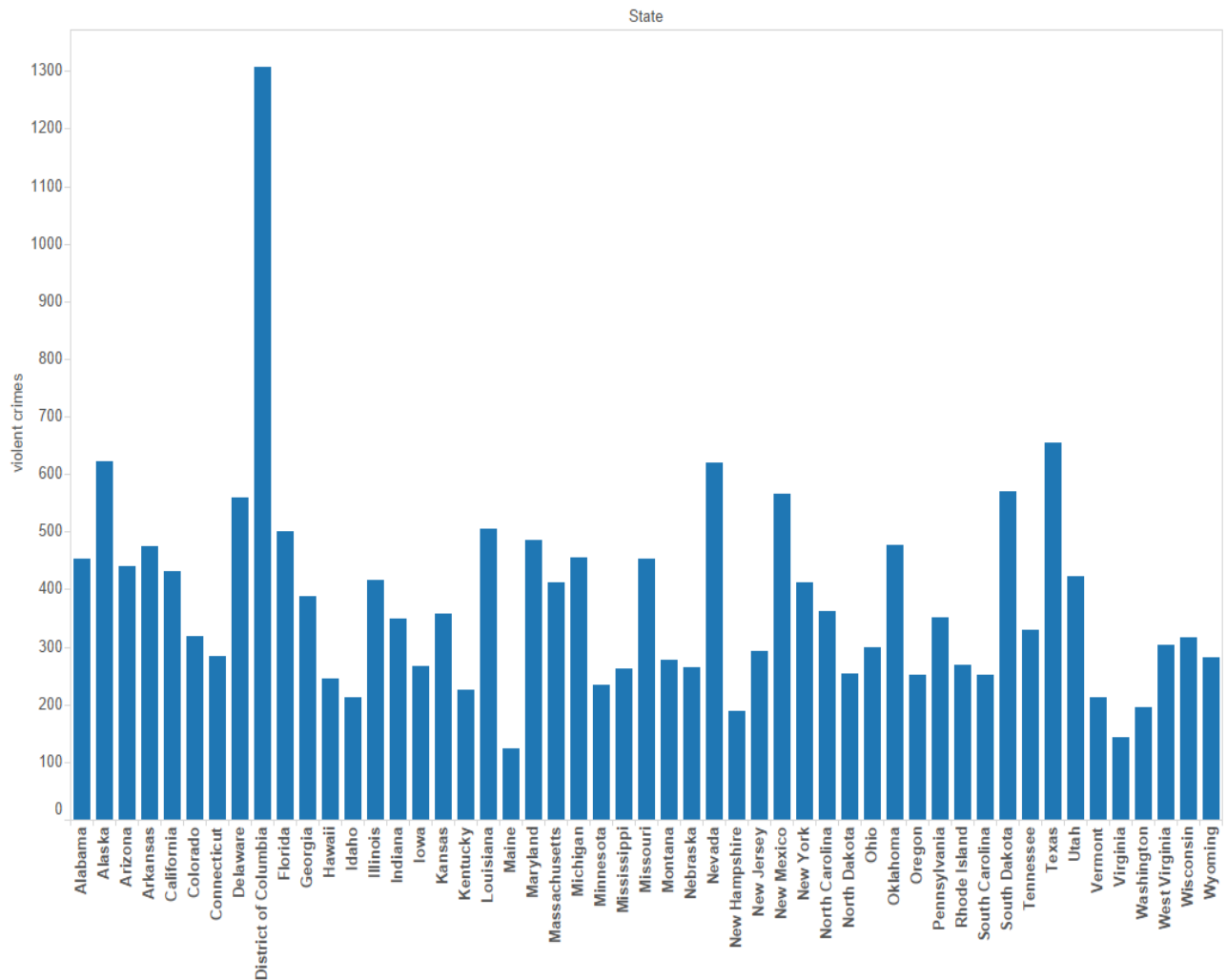
4.2 Histogram for violent crimes in 2011

Sheet 2

Sum of vcrime B for each State.

From the Histogram, it is evident that D.C has topped again the crimes for 2011, but the rate has come down drastically to 1234. Second highest rate is in Alaska rounding up to 617. Maine and Virginia are the least two respectively.

## 4.3 Histogram for violent crimes in 2012

Sheet 3

Sum of vcrime C for each State.

D.C still tops the list, but also the rates have risen back to 1306 almost similar to the rates of 2010. We could see Alaska has come down to 617 in the year 2010 from 638 in the year 2011 almost close to Texas which has a rate around 613.

4.4 Histogram for violent crimes in 2013

Sheet 4
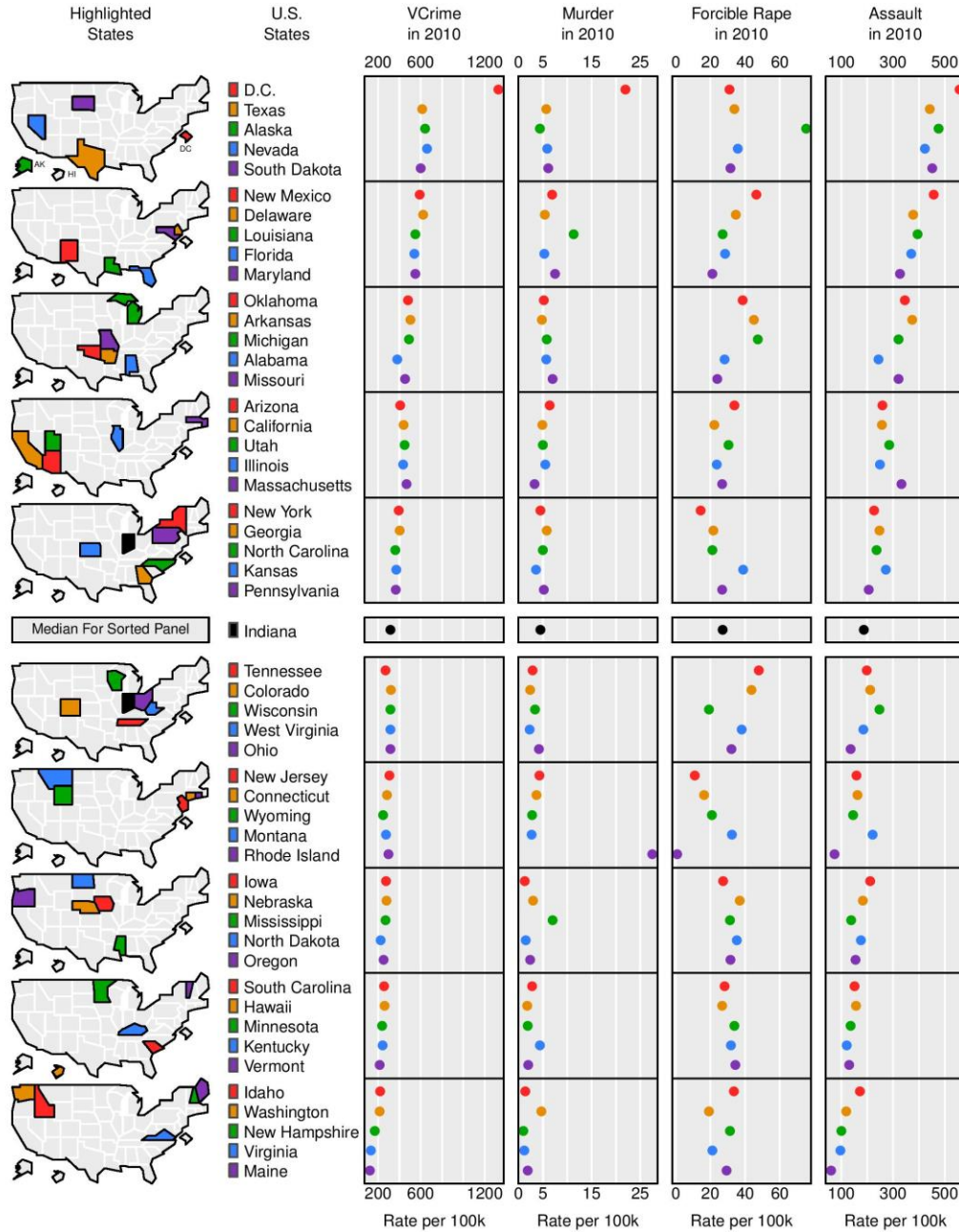

Sum of vcrime D for each State.
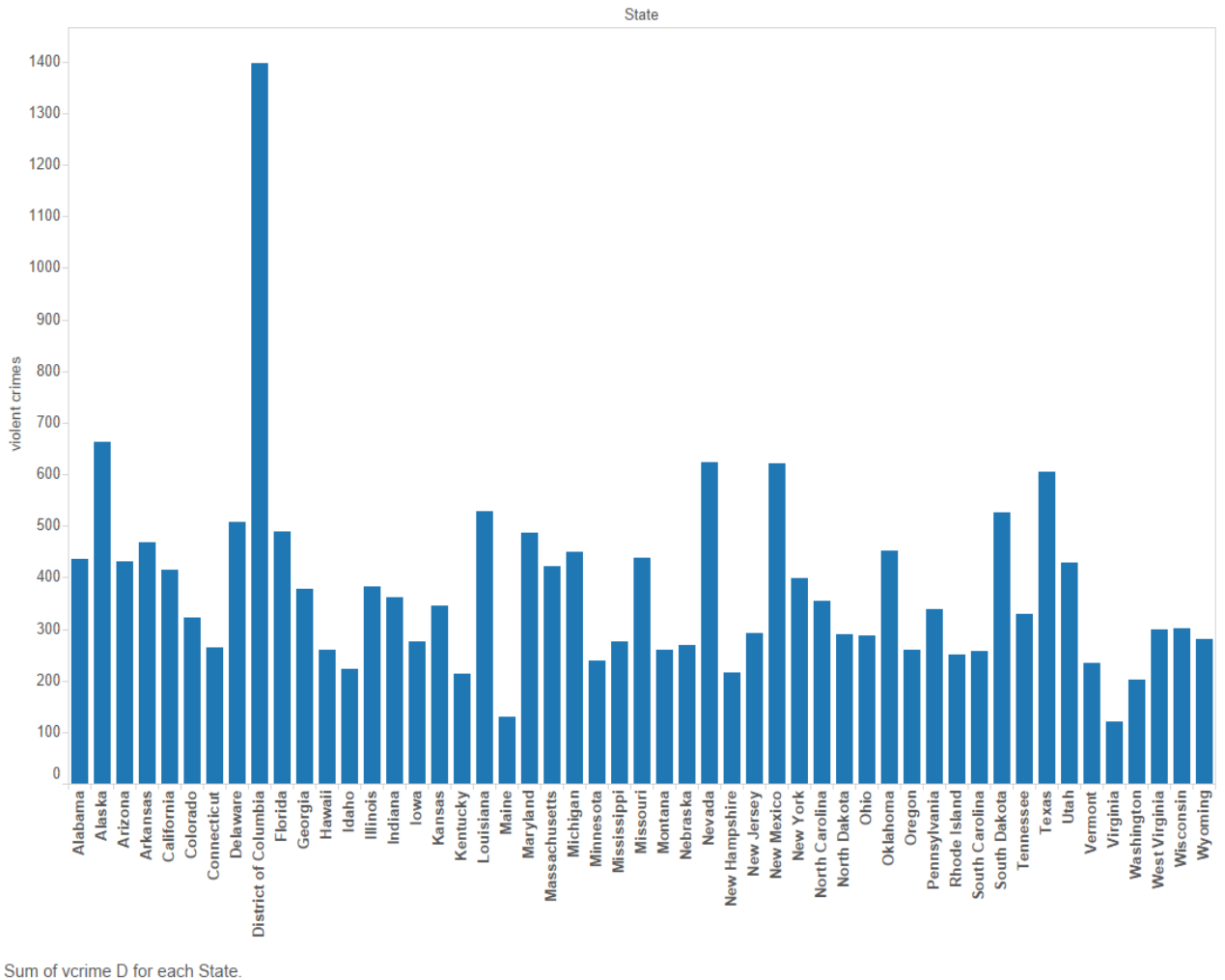
D.C interestingly still rises on crime rates, it's interesting to know what changed in 2011, for the rates to fall and rise again. Alaska also piked to 662 its 4 year high. Maine on the lowest crime in 2012 was beaten by Virginia rating 121, it's interesting to see what changes Virginia police devised.

5. Micromaps

The R function that I used to represent my dataset in the form of micromaps is micromapST. The data has been represented in 2 versions. One is using the dots other using bar plots. The type of cumulative shading used in representing the dataset is 'mapcum'. Mapcum is a type of cumulative shading from top to bottom which is used for spotting growing clusters of similar valued states from top to bottom across the panels. The glyphs used in the code are dot and bar.
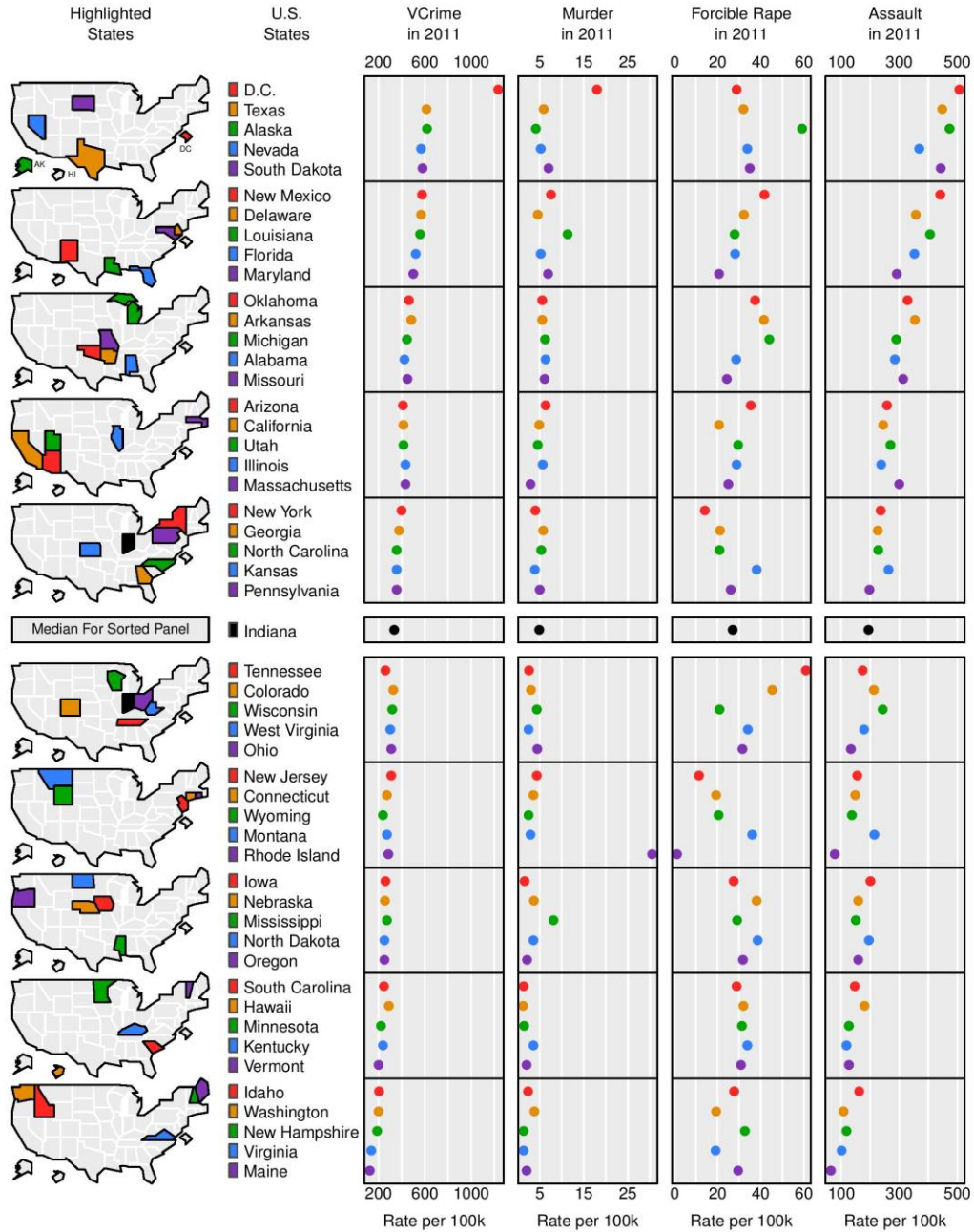
The following are the micromaps of crime rates in 2010, 2011, 2012, and 2013 respectively. Followed by the Micromap plot for Violent Crimes alone in these years.
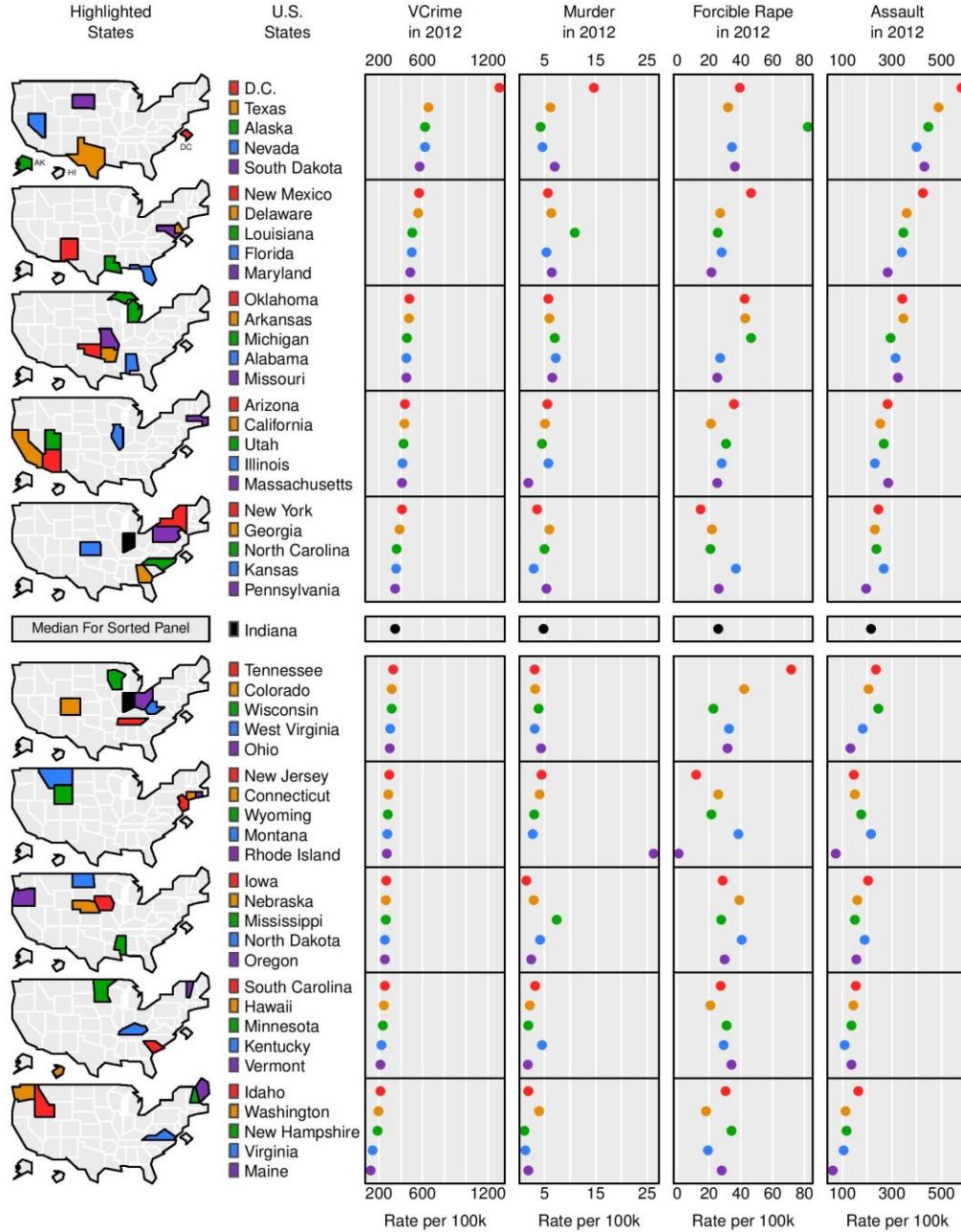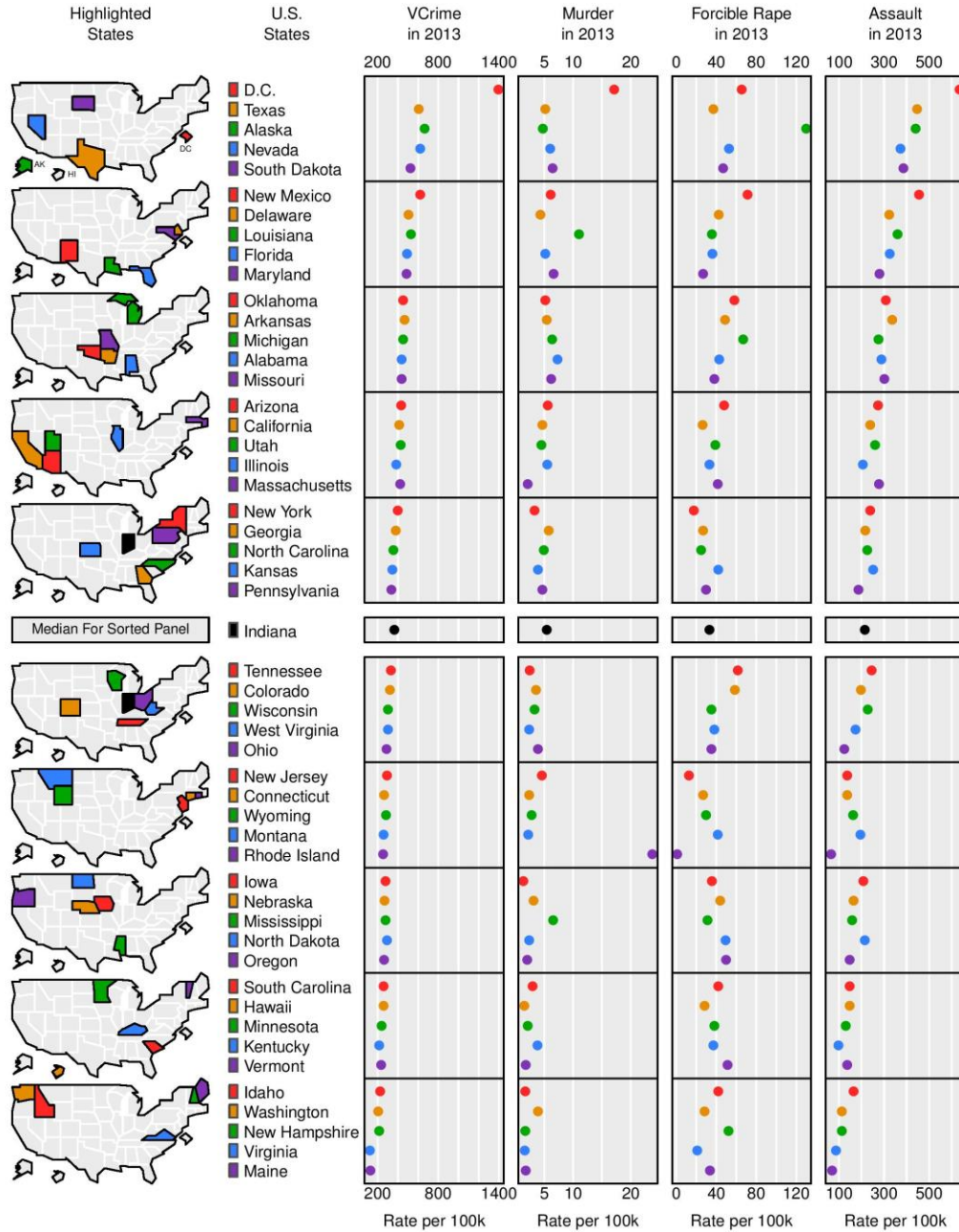
Violent Crimes
IN U.S.A

Highlighted States	U.S. States	VCrime in 2010	Murder in 2010	Forcible Rape in 2010	Assault in 2010

Rate per 100k

Violent Crimes
IN U.S.A

		VCrime in 2011	Murder in 2011	Forcible Rape in 2011	Assault in 2011

Highlighted States | U.S. States

D.C.
Texas
Alaska
Nevada
South Dakota

New Mexico
Delaware
Louisiana
Florida
Maryland

Oklahoma
Arkansas
Michigan
Alabama
Missouri

Arizona
California
Utah
Illinois
Massachusetts

New York
Georgia
North Carolina
Kansas
Pennsylvania

Median For Sorted Panel — Indiana

Tennessee
Colorado
Wisconsin
West Virginia
Ohio

New Jersey
Connecticut
Wyoming
Montana
Rhode Island

Iowa
Nebraska
Mississippi
North Dakota
Oregon

South Carolina
Hawaii
Minnesota
Kentucky
Vermont

Idaho
Washington
New Hampshire
Virginia
Maine

Rate per 100k | Rate per 100k | Rate per 100k | Rate per 100k

Violent Crimes
IN U.S.A

Highlighted States	U.S. States	VCrime in 2012	Murder in 2012	Forcible Rape in 2012	Assault in 2012
		Rate per 100k	Rate per 100k	Rate per 100k	Rate per 100k

Violent Crimes
IN U.S.A

Highlighted States	U.S. States	VCrime in 2013	Murder in 2013	Forcible Rape in 2013	Assault in 2013

Rate per 100k | Rate per 100k | Rate per 100k | Rate per 100k

Median For Sorted Panel — Indiana

States listed (in colored groups of five):

- D.C., Texas, Alaska, Nevada, South Dakota
- New Mexico, Delaware, Louisiana, Florida, Maryland
- Oklahoma, Arkansas, Michigan, Alabama, Missouri
- Arizona, California, Utah, Illinois, Massachusetts
- New York, Georgia, North Carolina, Kansas, Pennsylvania
- Tennessee, Colorado, Wisconsin, West Virginia, Ohio
- New Jersey, Connecticut, Wyoming, Montana, Rhode Island
- Iowa, Nebraska, Mississippi, North Dakota, Oregon
- South Carolina, Hawaii, Minnesota, Kentucky, Vermont
- Idaho, Washington, New Hampshire, Virginia, Maine

Violent Crimes
IN U.S.A

	Highlighted States	U.S. States	VCrime in 2010	VCrime in 2011	VCrime in 2012	VCrime in 2013

Rate per 100k   Rate per 100k   Rate per 100k   Rate per 100k

Interestingly, we could see that how each state could change its Crime rates for the years 2010-2013. It is visually very appealing to use a micromap to plot the data and observe the changes in a glimpse. We could observe the changes in the scale of 2011 from 1200 to 1000, due to the drop of crimes in D.C. And also states like Colorado, Hawaii and Tennessee's changes across the years are visually evident. Another advantage of Micromap is Indiana always was in the median region, this median display helps statisticians to use this as a reference point for further analysis on data.

5.6 Cumulative Micromap for Violent Crimes in 2010

Cumulative Micromaps are a way to see how the crime rates are being totaled from the highest rate to lowest rate by state order. It is an easy way for analysts to see how far the combined effects on a specific region, for eg: in West or North. It would also help in further analyze and see if other factors specific to a particular region play any role in the results.
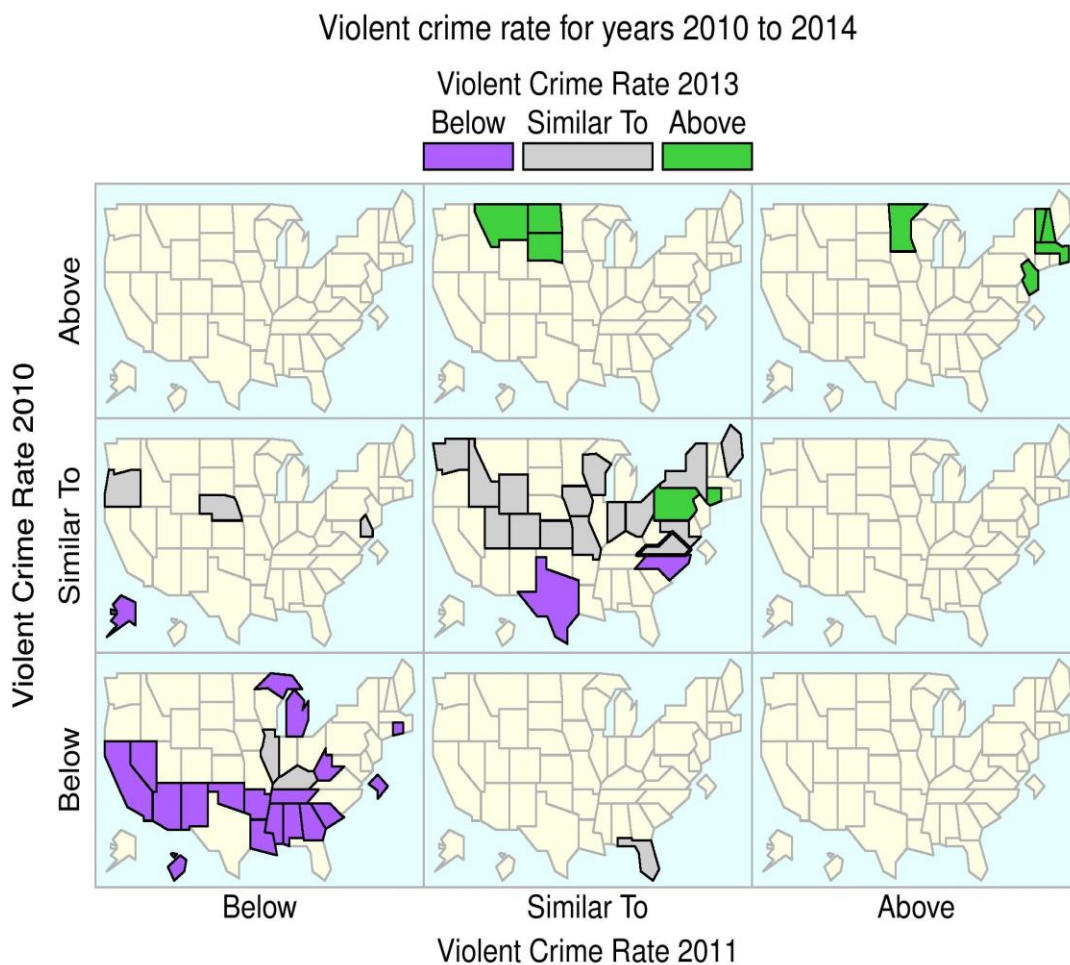
VCRIME RATES
IN U.S.A
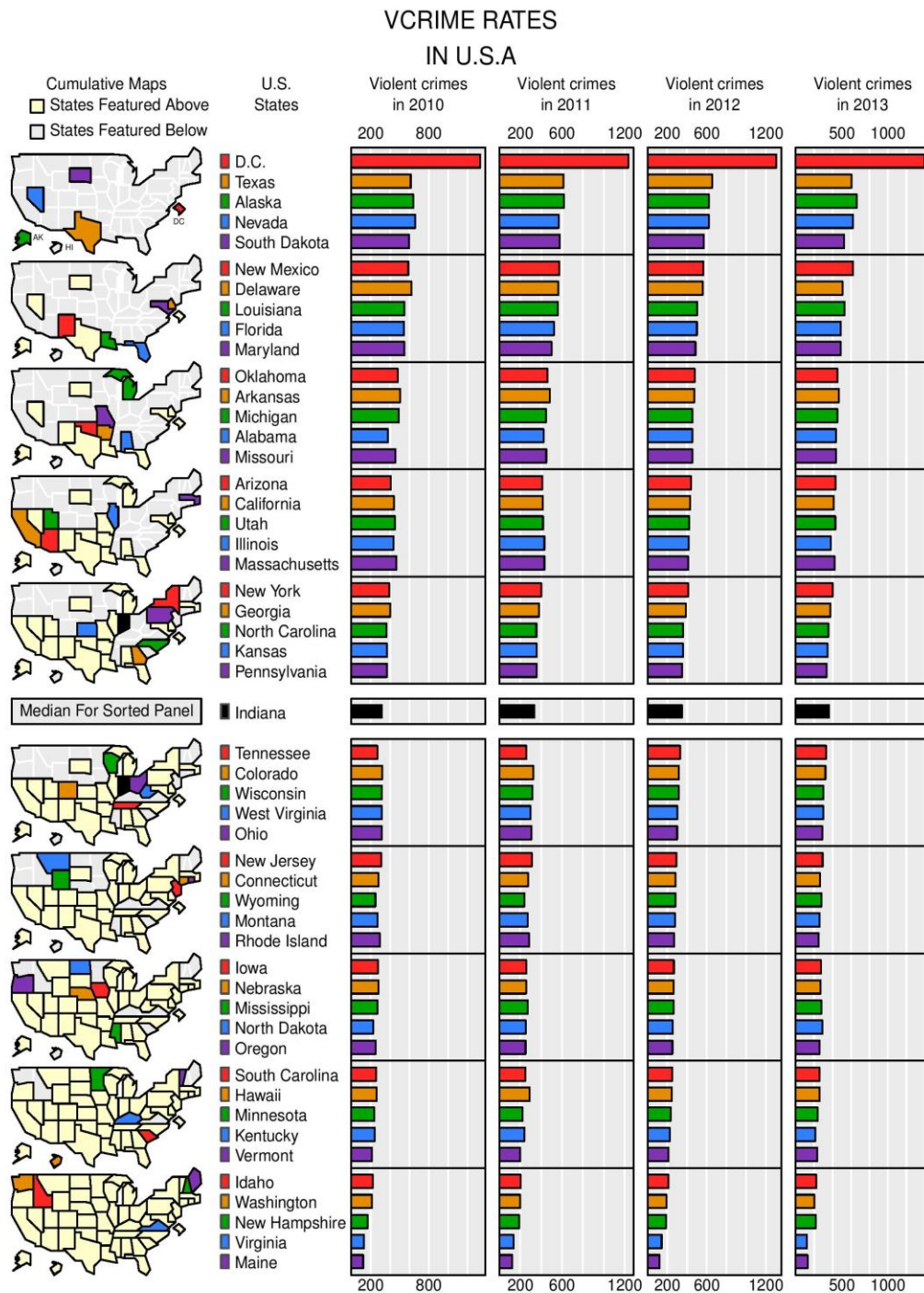
4.7 Tailed Micro map for violent crimes in 2010

5.7 Tailed Micro map for violent crimes in 2010

Violent CRIME RATES IN U.S.A
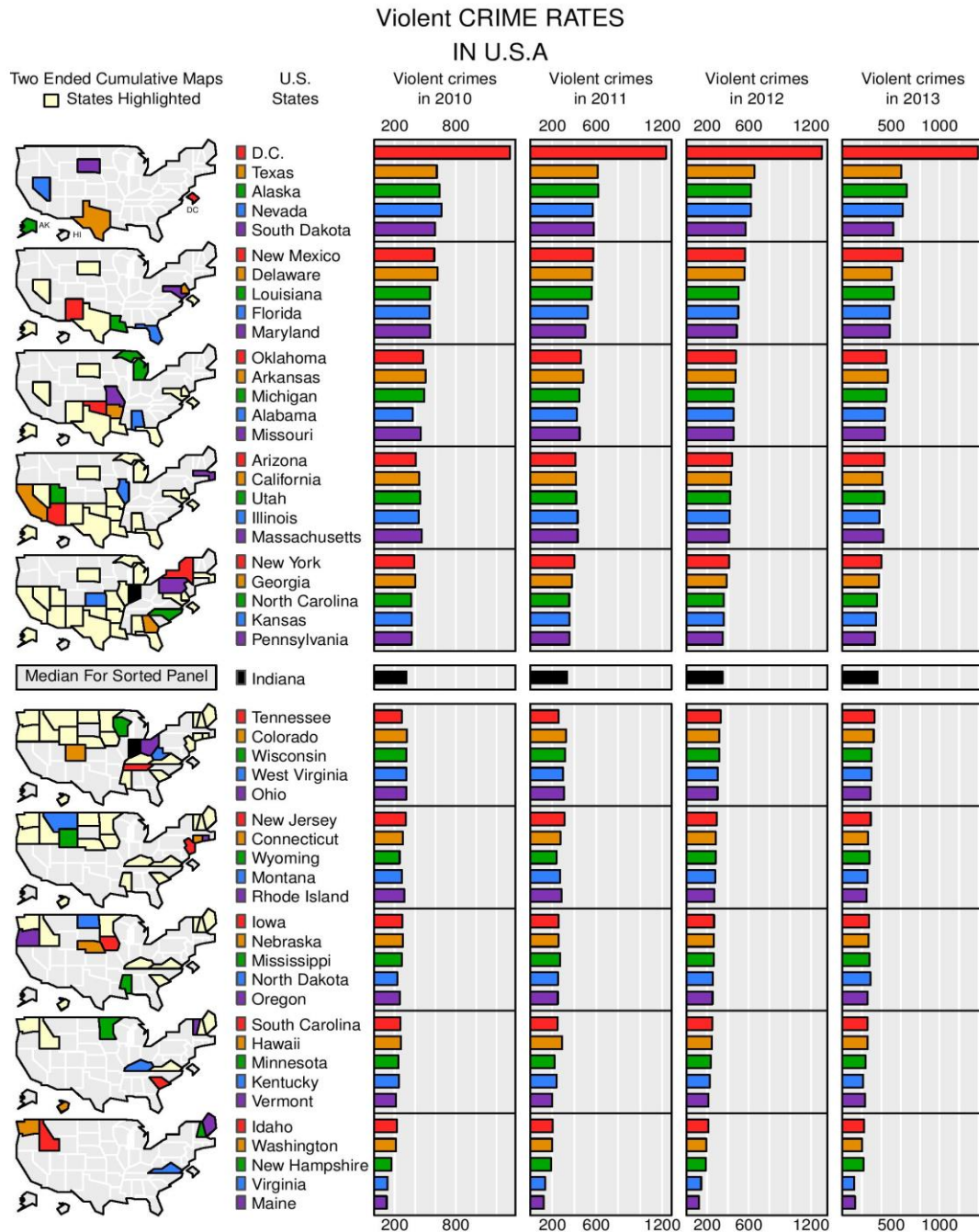
I personally feel tailed Micromap provides an interesting perspective in representing the data. In the top half of the plot we can see how the states with highest rates cumulate till the median, and the lowest rate states cumulate upwards till the median. When we combine the upper and lower plot of median we obtain the complete plot of states. This is a wonderful analysis, which helps us first sort the data like a micromap, second gives a cumulative, and third helps us also with a new perspective of analysis.

6. Confident Class Map:


Violent crime rate for years 2010 to 2014

The Violent crime rates from the years 2010 to 2013 have been divided into three class intervals in a CC map, with interval ranges of:

- Crime rate between 0-250, is Ranked 1

- Crime rate between 250-450, is Ranked 2

- Crime rate greater than 450, is Ranked 3

The Confident class map shows three different intervals in three distinct colors and represents the ranks by these colors. Below (Rank 1) is Violet, Similar to (Rank 2) is Grey and Above (Rank 3) is Green. Thus, CC maps help when we do not have a data set and would like to figure out whther a specific state falls under or above a certain threshold. Though, we can't get the exact figure, we can get a vague idea on where a state lies in its crime rate.

7. Linear Regression:

Linear Regression of Crime rate to Law Enforcement Employees in the year 2011:
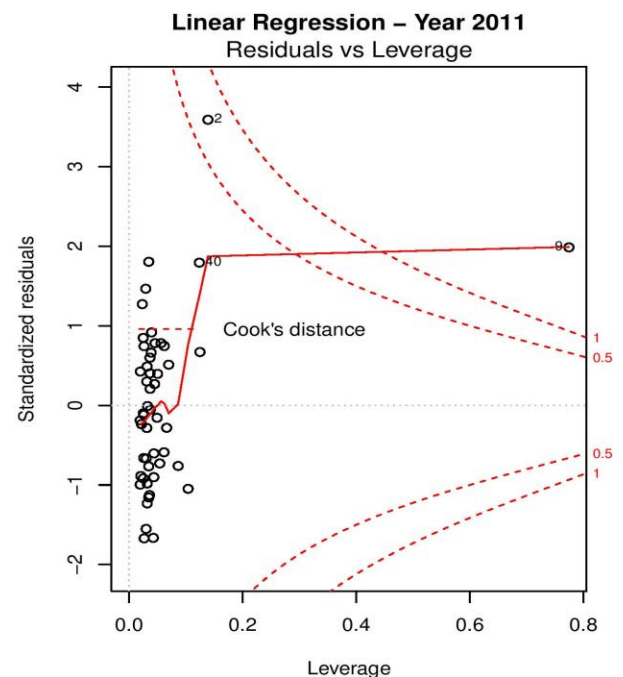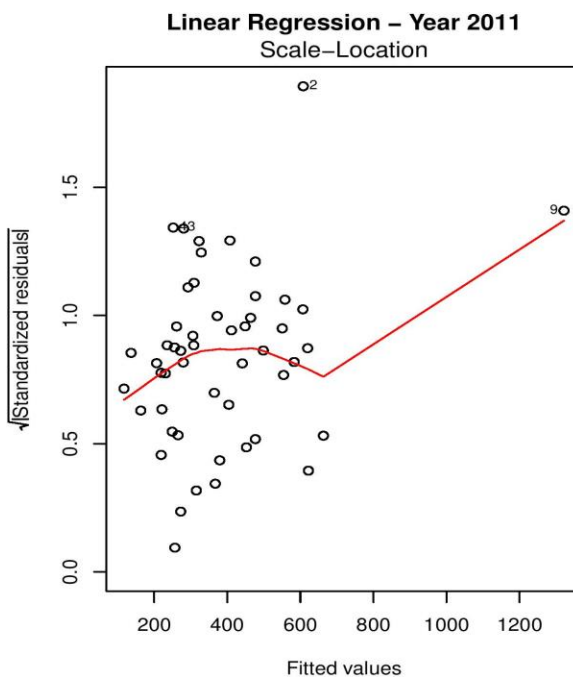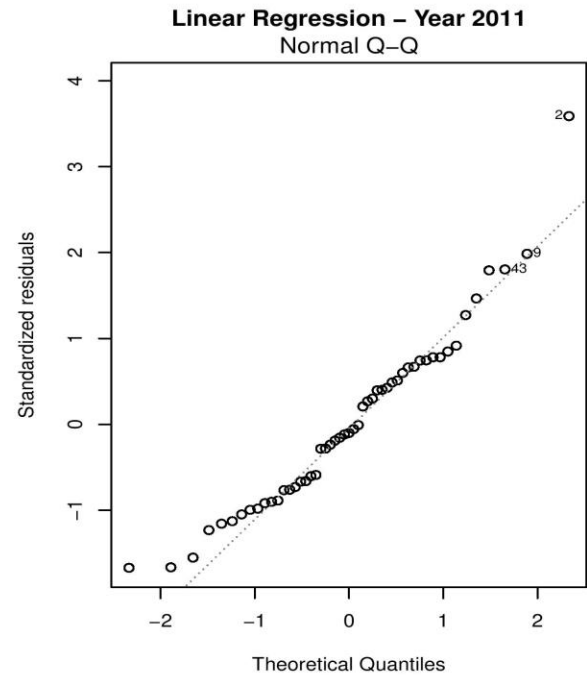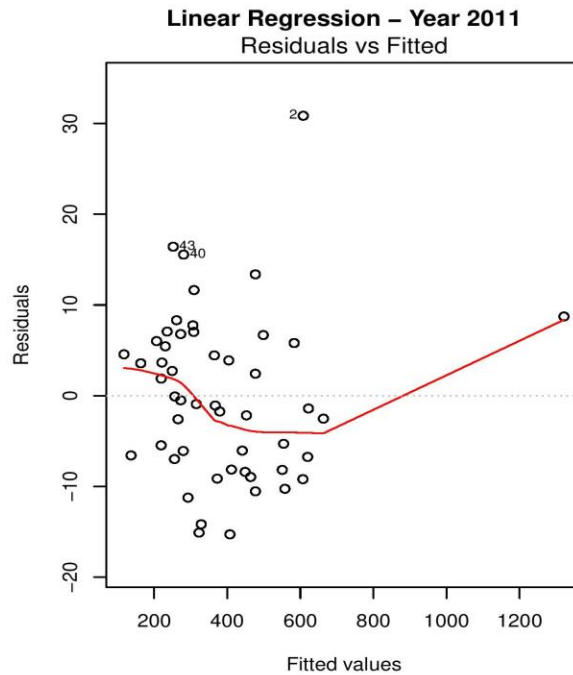
R Squared value – **0.9977**

P value – **2.2e-16**

I have taken **Regressand** (Base value) as **Violent Crime rate** in 2011 and **Regressor** as **Law Enforcement Employee rate** in year 2011. I tried to map if any trend exists between these two entities. Interestingly the presence of police has increased crimes, which was not what I thought initially. That would be a great subject to further research on.

Residual Vs Fitted

This plot shows the residuals (the vertical distance from point to the regression line) versus the fitted values. The red line which is smooth curve is very near to the gray dashed line. So this explains that there is more correlation between the two variables, Crime Rate and Law Enforcement employee Rate.

Normal Q-Q

This plot explains about the errors. In the Normal QQ plot, we can see that the most of the points lie on the grey line, so this mean that the errors are normally distributed.

Scale – Location

In this plot we can see that there is somewhat equal statistical variances because the red line is looking horizontal.

Residuals Vs League
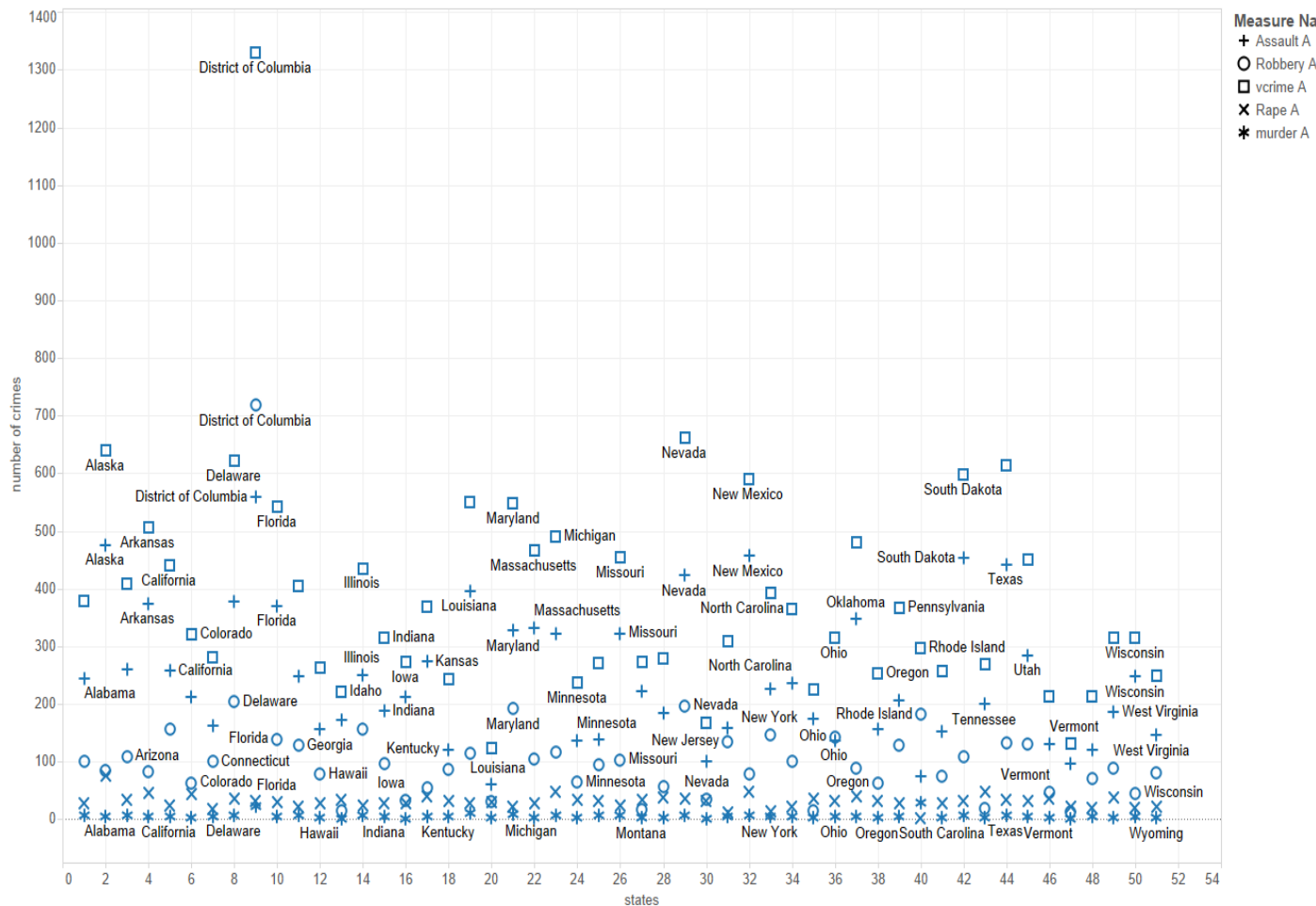
Standard residuals are not centered at Zero.

Cook distance is not greater than 1.0

8. Scatterplots:

Using Tableau, when scripted for a scatter plot for the different types of crimes in 2010 to 2013 the result are as follow:
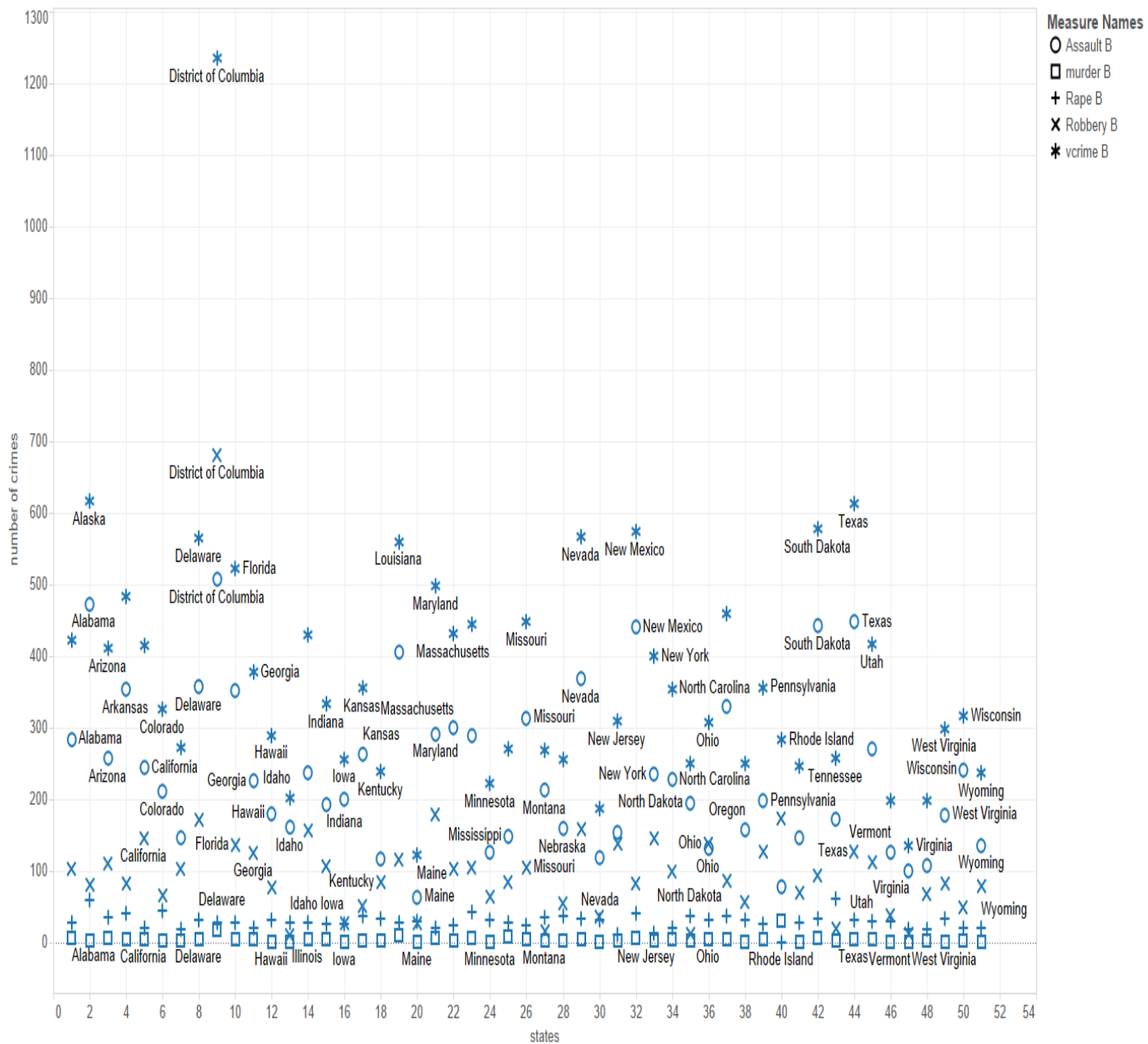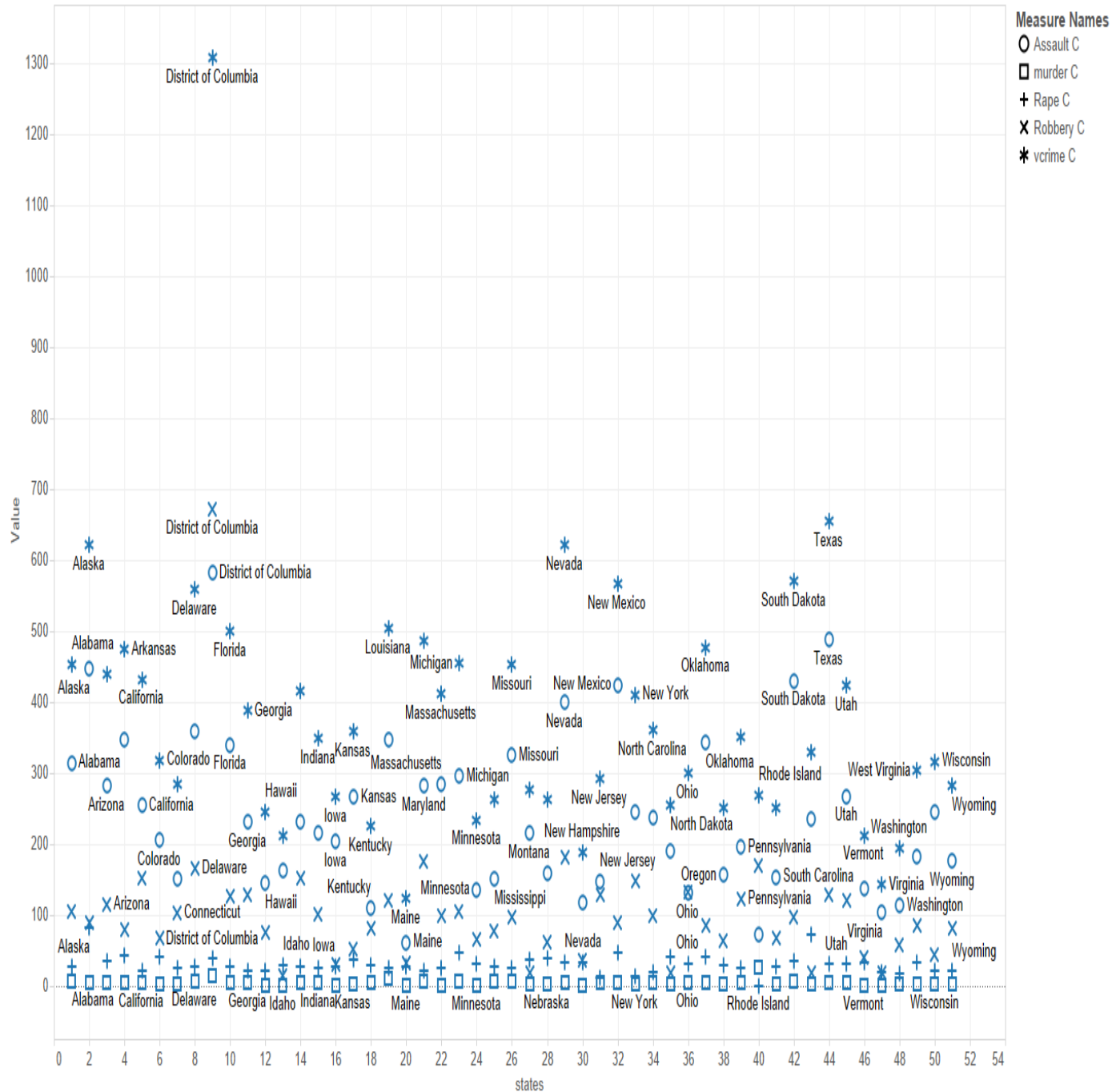
8.1 Scatterplot for year 2010:

Sheet 3

Sum of Order vs. Assault A, Robbery A, vcrime A, Rape A and murder A. Shape shows details about Assault A, Robbery A, vcrime A, Rape A and murder A. The marks are labeled by State.

8.2 Scatterplot for year 2011:

Sheet 6

Sum of Order vs. Assault B, murder B, Rape B, Robbery B and vcrime B. Shape shows details about Assault B, murder B, Rape B, Robbery B and vcrime B. The marks are labeled by State.
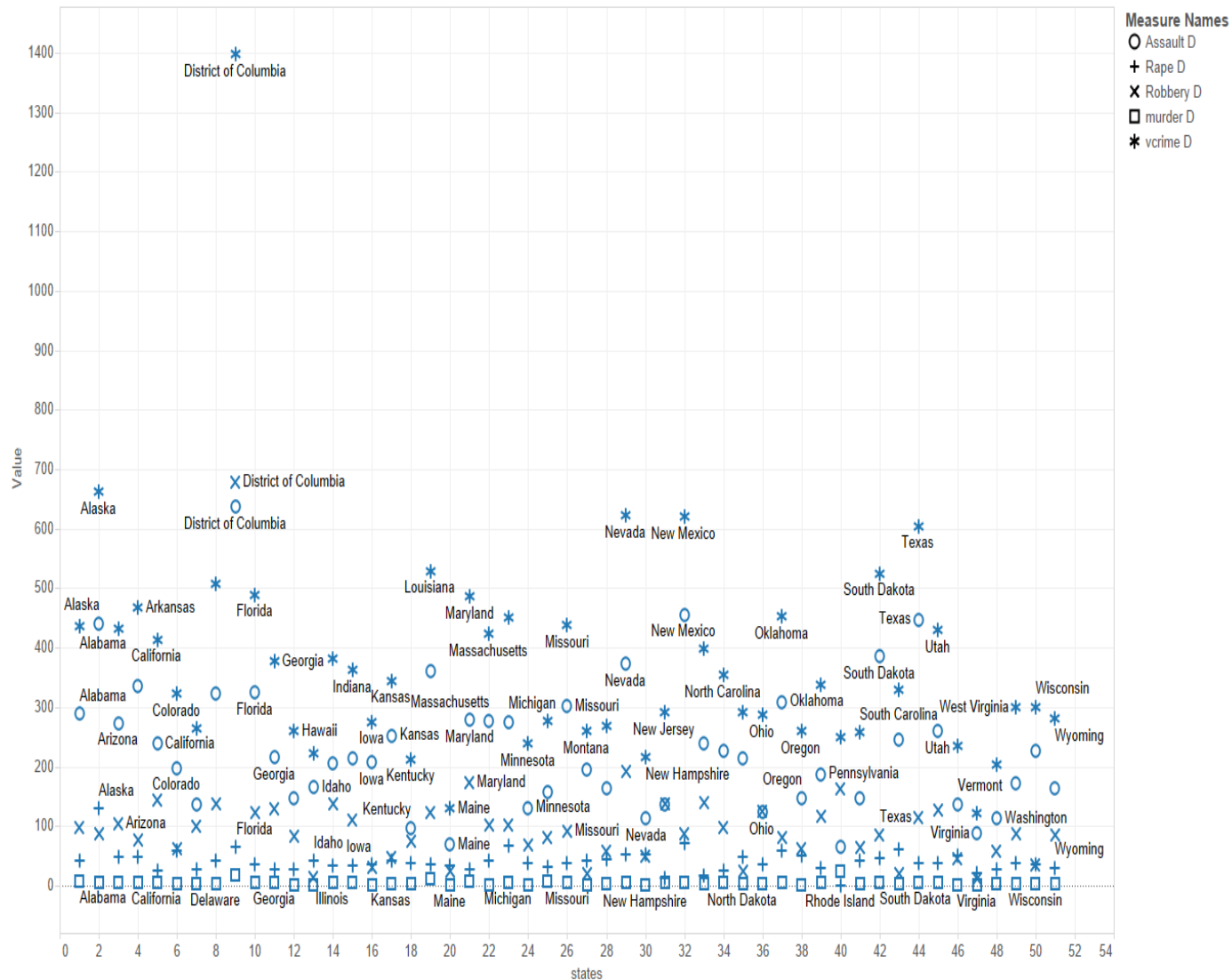
8.3 Scatterplot for year 2012:

Sheet 6

Sum of Order vs. Assault C, murder C, Rape C, Robbery C and vcrime C. Shape shows details about Assault C, murder C, Rape C, Robbery C and vcrime C. The marks are labeled by State.

# 8.4 Scatterplot for year 2013:

Sheet 6

Sum of Order vs. Assault D, Rape D, Robbery D, murder D and vcrime D. Shape shows details about Assault D, Rape D, Robbery D, murder D and vcrime D. The marks are labeled by State.

Scatter plots offer a variety of options like percentage changes, analysis of all the entities of a huge data. Also it represents the different attributes for the state, along vertically. This provides an analysis on if different attributes of a state are interlinked.

9. TC Maps:

With the normalized data, the violent crimes for 2010, 2011, 2012, and 2013 have been plotted.

9.1 Cognestic View:

Cognostic View
R-Squared = 63.39 %
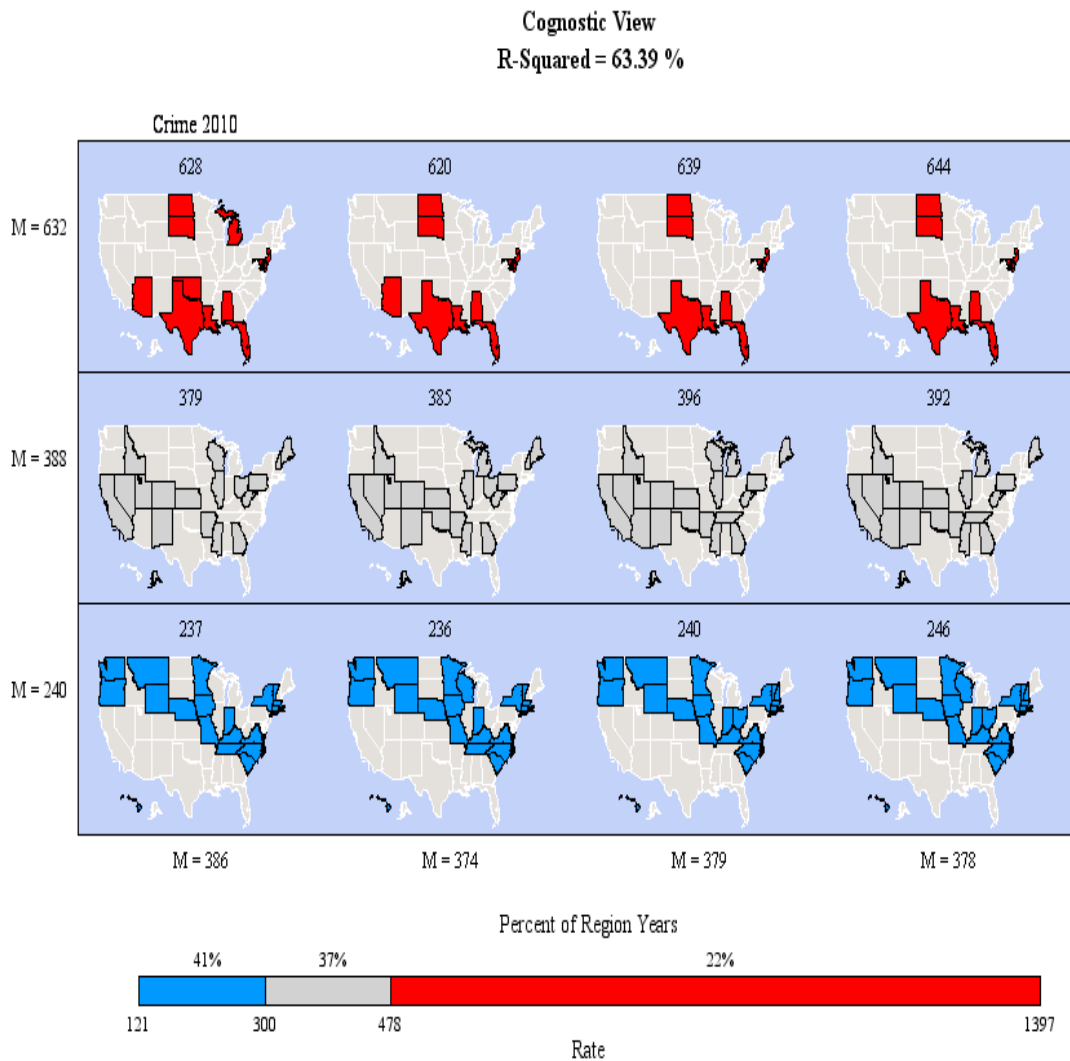
The plot is divided into three regions with minimum value of 240, median of 388 and a maximum of 632. From the graph it is evident that 41% of the total states have crime rates around 240, 37% have around 388 and 22 % have crime rates around 632. The R squared (coefficient of determination) = 63.39%.

9.2   Change Series Map:
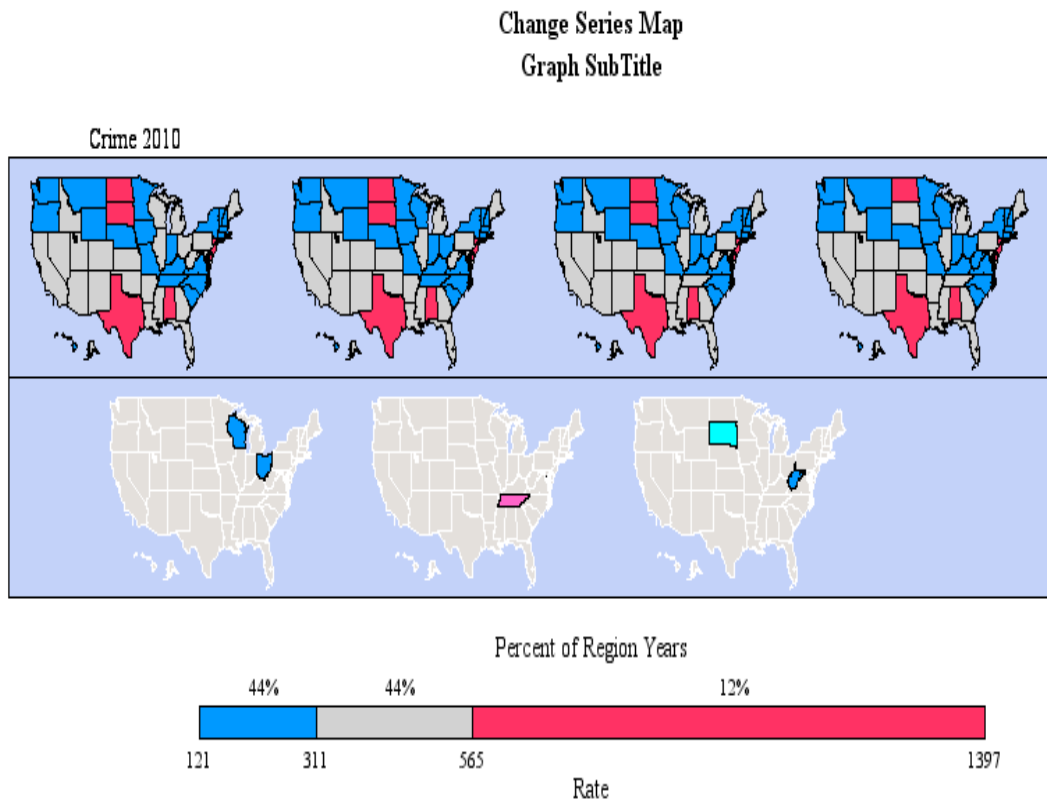
This map shows the states with Violent Crimes for the four years in the first row. The second row shows the changes of the Violent crime numbers from 2010 to 2011, 2011 to 2012 and 2012 to 2013 respectively. We can see there is significant change between the years 2010 & 2011 and not much change in between the years 2012 & 2013, and some change in 2012 & 2013.

9.3   Rising and Falling Map:
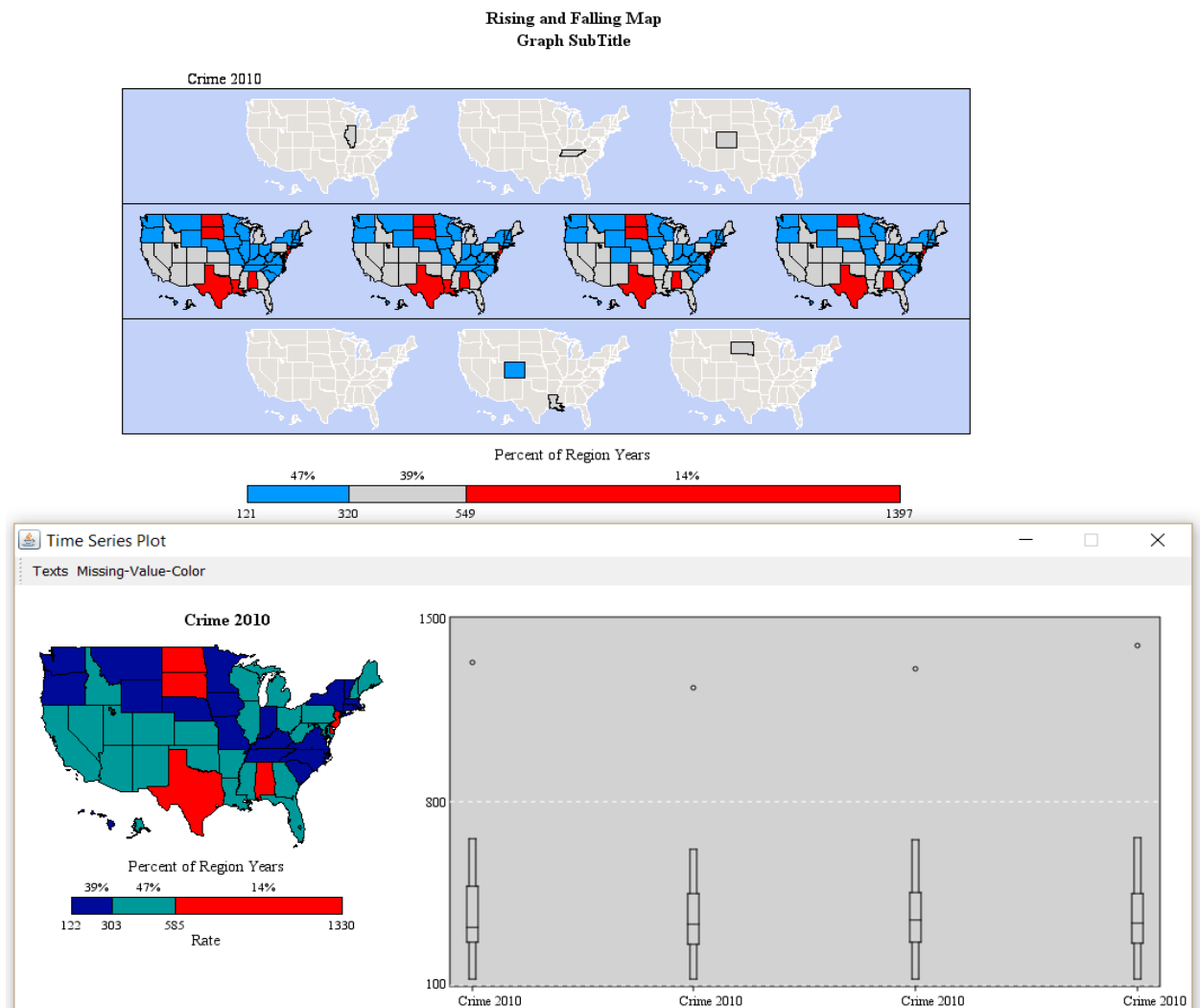
This plot displays the rise and fall of the total violent crime numbers between the years of 2010-2013 for all the states.

10. Conclusion:

Datasets represented graphically are more appealing than the data in the row-column form. For representing the data I have used TC maps, Scatterplot, Histograms, Linear Regression, Confident Class and Micromaps.

On observing the various represented maps, it is evident that D.C topped the list of crime rates for all the years followed by Alaska, and Virginia ended up being the least by beating Maine.

We could see how rates dropped in various states in certain years. And, how they fluctuated between the observed timeline. Though Violent Crime, Murder, Rape, Robbery, and Assault are analyzed, much emphasis is kept on Violent Crimes for further analysis.

Also, a linear regression model was used to check if any trend exists between the number of Law Enforcement Employees and Violent Crimes per state. It was interesting to see that Arkansas, Mississippi, New Mexico and Texas have 1:100 ratio, which signifies the policing in the state should be increased and this might mean a good way to reduce Violent Crimes.

On the contrary in states like California and Idaho though the ratio is small 1:10 & 1:4, the crime rates don't seem to be affected by police presence, which would contradict our previous analysis of better policing would mean safer place to live.

I believe this would be a great field of interest for further research. I am planning to continue this work by trying out analysis using different variables like how different geographical locations, latitude, G.D.P, diversity and psychological behavior patterns of a region affect crime rates.

REFERENCES:

1. www.data.gov.

2. http://www.statnetconsulting.com/micromaps.html.

3. R Studio Desktop. Vers. 0.98.1103. N.p., n.d. Computer software.

4. Stat 515 course assignment scripts, BlackBoard, GMU.