Campus Safety and Security Process Book

Mohsen Abbasi*

Amir Biglari†

Waiming Tai‡

1   Overview and Motivation

Choosing an institution in USA and even choosing an appropriate US state or city is a major decision for students and their families. Along with academic, financial and geographic considerations, the safety of campus is a vital concern. This work tries to investigate the campus safety by considering on-campus crime records all over the US.

Students can better compare the crime records in US States and universities using the visualizations in this work. In addition, social scientists can use this visualized data to find in what universities and states there are more problems related to crimes. They can also see the record of different types and categories of crimes separately at different schools and location. This may help them find if there is any relation between an specific crime type/category to its geographic location.

2   Questions to answer

By visualizing this data set, we have a few questions to answer and objectives in mind:

- Answering basic questions such as:

  1. Which states have the safest campuses across the United States?
  2. Find the safest and unsafest schools campuses across the United States.
  3. Find the safest and unsafest schools campuses across a specific state.
  4. Which schools are facing some specific type of criminal activities more than the others?

*UID: u1011952, Email: m.abbasi@utah.edu

†UID: u0613945, Email: amir.biglari@utah.edu

‡UID: u1008421, Email: u1008421@utah.edu

5. Which schools are similar considering a particular type of crime?

6. Is there any relation between a specific type of criminal activity on-campuses with their geographic locations?

7. Which type of crime has been the most threat to the students at a specific year at different locations?

8. How does the crime rate has changed in time?

- We're also interested in comparing different schools in details and finding out what types of crimes plays a more significant role in threatening the safety of students of a school.

- Categories such as criminal offenses, disciplinary actions, hate crimes and violence against women are included in the data set. As a future work or continuation of this work, it will be interesting to combine this data set with other ones containing demographic information, GDP and such for different states and look for correlations between the crime rates and these attributes.

3 Data

The datasets that we worked with is about "The Campus Safety and Security". The raw data for this work can be accessed from "http://ope.ed.gov/campussafety/#/". This data set is so inclusive and contains the information about all the universities across the US with all of their campuses inside or outside the US, all institution sectors, and all programs. This data can also be sliced according to different school sizes. This database also includes information about crimes related to on-campus, on-campus housing, non-campus, public properties, and Reported by locals and state police. It has the records related to the years between 2001 and 2014.

However, in order to reduce the volume of the data and also considering the more important, and more popular schools which are more appropriate for our works, we filter the data before downloading it from the website.

For this work, we only considered the crimes conducted on-campus between 2001 and 2014. We filtered down the institution sectors to 1) Public, 4-year or above, 2) Private nonprofit, 4-year or above, ans 3)Private for-profit, 4-year or above. We also filtered the data with respect to institution enrollment (that shows the institution size) and only considered the schools with more that 10000 students.

The crimes covered in the data set are categorized into criminal offenses such as theft, disciplinary actions such as criminal actions, hate crimes, VAWA offenses and others. Disciplinary actions has categories of illegal weapons possession, drug law violations, and liquor law violations. Criminal offenses includes murder/non-negligent manslaughter, negligent manslaughter, sex offenses-forcible, rape, fondling, sex offenses-nonforcible, incest, statutory rape, robbery, aggravated assault, burglary, motor vehicle theft, and arson. Hate crimes has 150 sub-categories and violent against women includes domestic violence, dating violence, stalking.

Fortunately (for the sake of safety on campuses), "The Campus Safety and Security" data set is a sparse one (there are many zero values in each record). We considered all of those entries as 0. Also, note that the record related to violence against women has only conducted at 2014. We copied this record to all the previous years for better visualization.

The database can provide all the data for universities together, or separated with respect to states. But it won't generate a data that includes state tags. Therefore, for data related to the states, we needed to process the data and combine some of it. Also for each of the four categories, we needed to calculated their sum from their subcategories.

4 Design Evolution and Related Works

There are three phenomenons we would like to study and visualize:

1. The crime statistics for different institutes in each state. Also, see how these statistics have changed over time.

2. Analyze the statistics by choosing a specific type of crime. For each type of crime, the trend for different categories of that crime over time will be shown in graphs.

3. Comparing the crime statistics between different schools. By choosing two or more schools, different types of crime would be compared.

The following are the designs we'd like to consider:

4.1 Must-Have and Optional Features:

We should be able to easily observe the crime statistics for different schools in each state and be able compare them. We should be able to see how different states are compared to each other in terms of crime statistics in their schools. Also, we should be able to see how these statistics have changed over the years.
As mentioned in the objectives, we'd like to combine this data set with other ones containing demographic information, GDP and such for different states and look for correlations between the crime rates and these attributes.

4.2 Crime statistics in each state:

This design would be to display the map of the United States. Here each state is colored using a gradient color map that uses saturation with respect to the crime number per students in that state. The map is broken down by states where each one is selectable. There is also a time line bar that can be adjusted to the desired year. Beside the map there would be a line chart showing the trend of the change in the number of different types of crimes in US universities in time (Fig. 1). By selecting a specific type you would be able to see a bar plot showing the amount of crimes for different crime categories of that type (Fig. 2). By hovering over a state you can see a general information about it (Fig. 3). By choosing a state, line chart and the barchart beside the map would get updated with respect to that

specific state. The general crime statistics would also be displayed below the map on a line chart in which each line represents a school in the selected state (Fig. 4) and by hovering over the lines you can see the schools names (Fig. 5). Also, when a line in the line chart is selected, the detail breakdown of different types of crimes would be displayed on a stack bar chart (Fig. 6). By selecting each chunk of the stack bar chart, you will get another stack bar chart for the categories of that crime type for the specific school (Fig. 7). There would be two search boxes to search any two schools and compare their crime statistics. By selecting a pair of schools and do hitting the compare button the user will get a new view which compares the school crime statistics in line charts and bar charts.
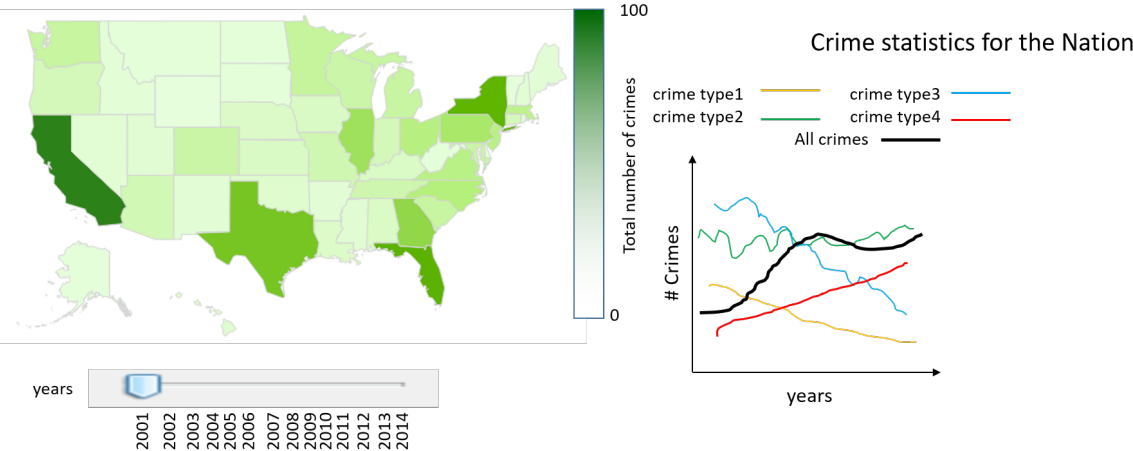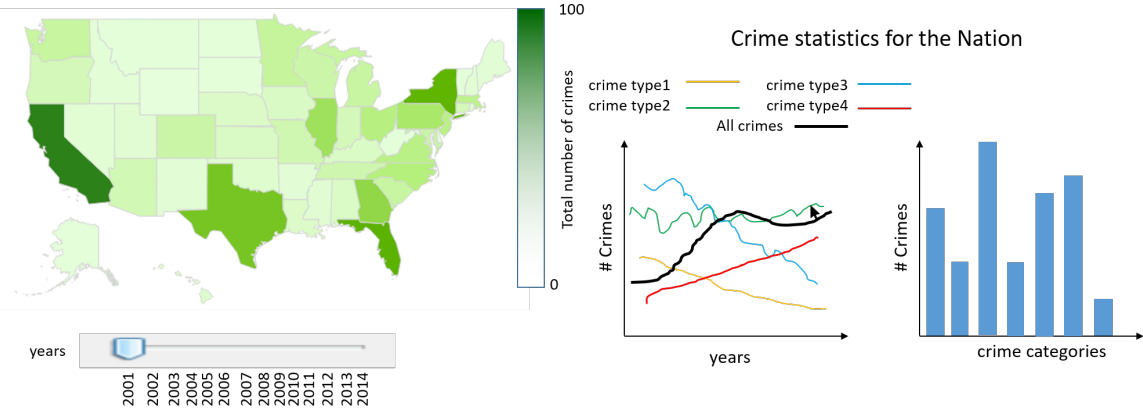

Figure 1:

Figure 2:

4.3   Clustering the data

This design will also have the map of the United States. This time we use a scatter plot chart to visualize the data for the whole nation or the selected states (Fig. 8). In this scatter plot,
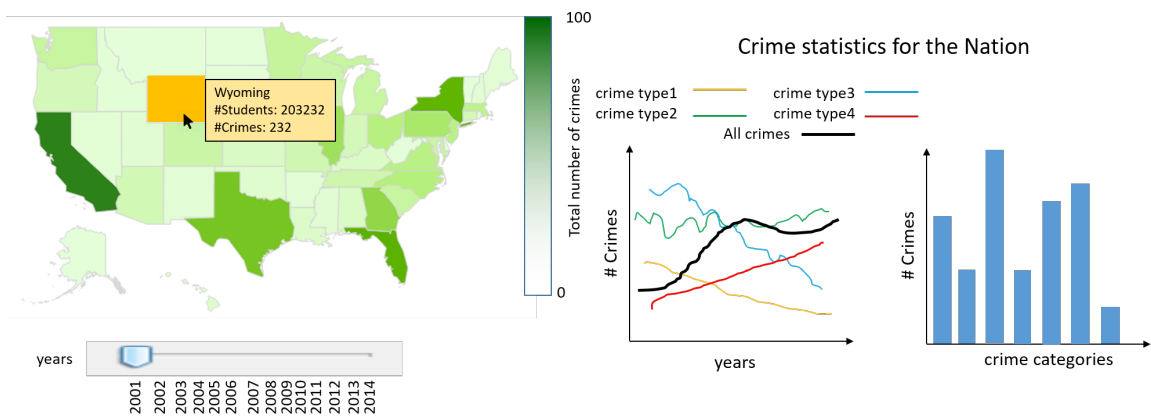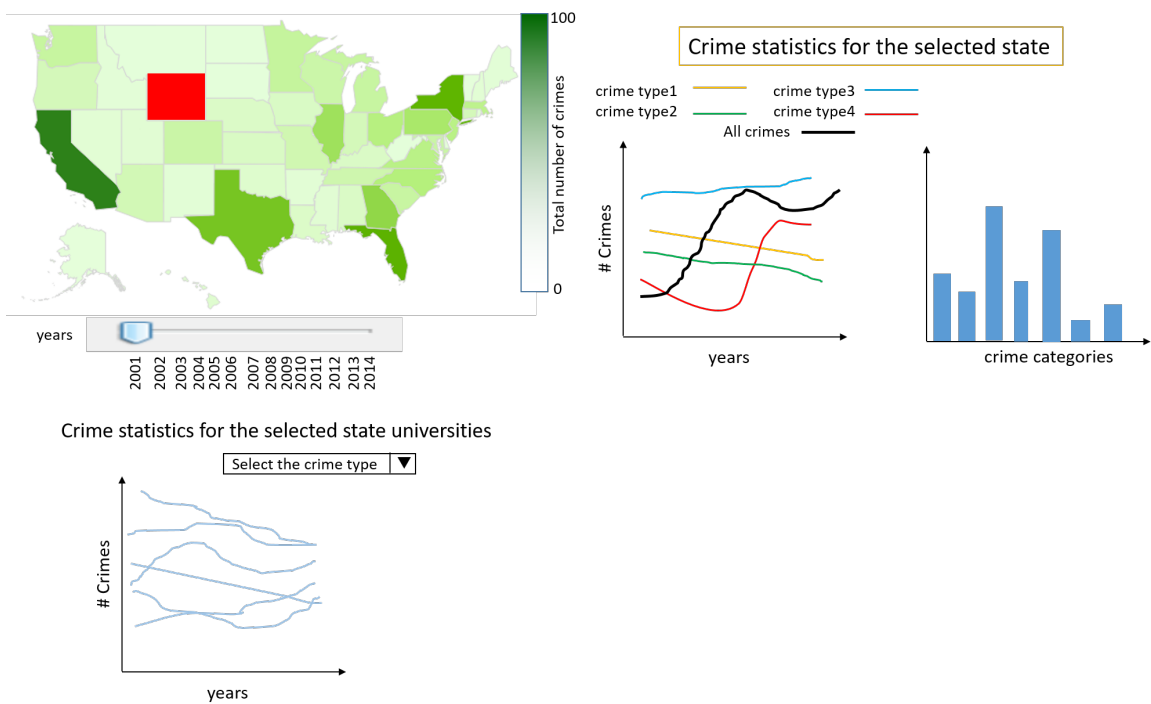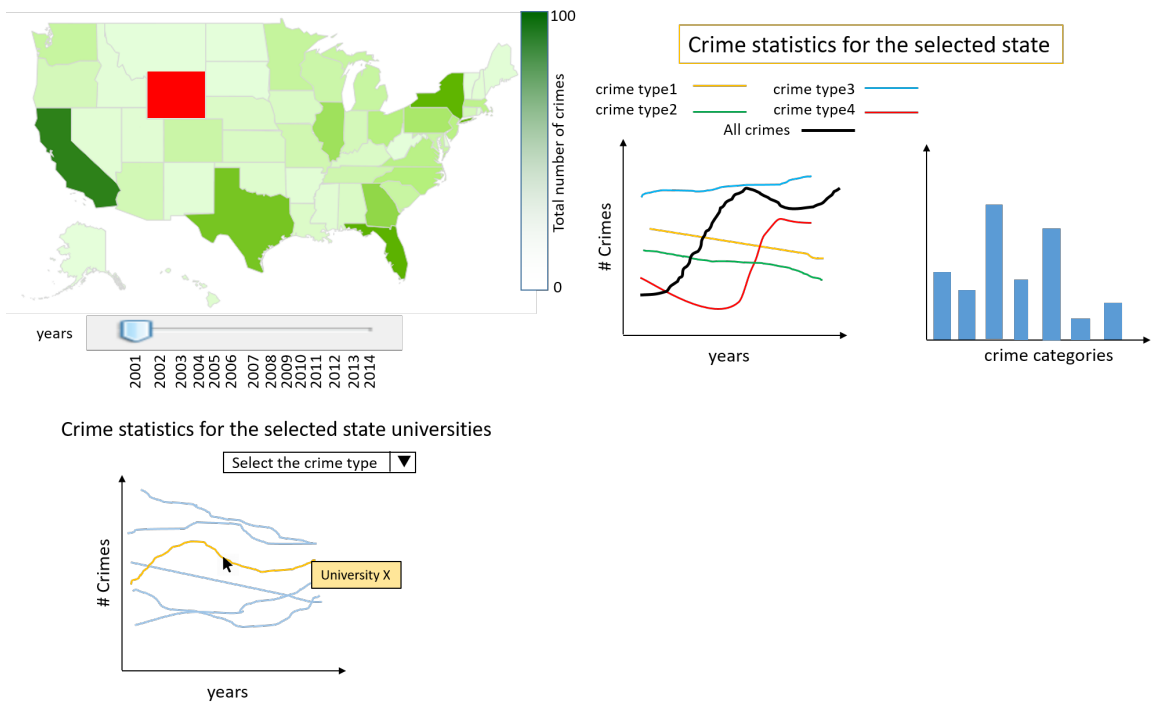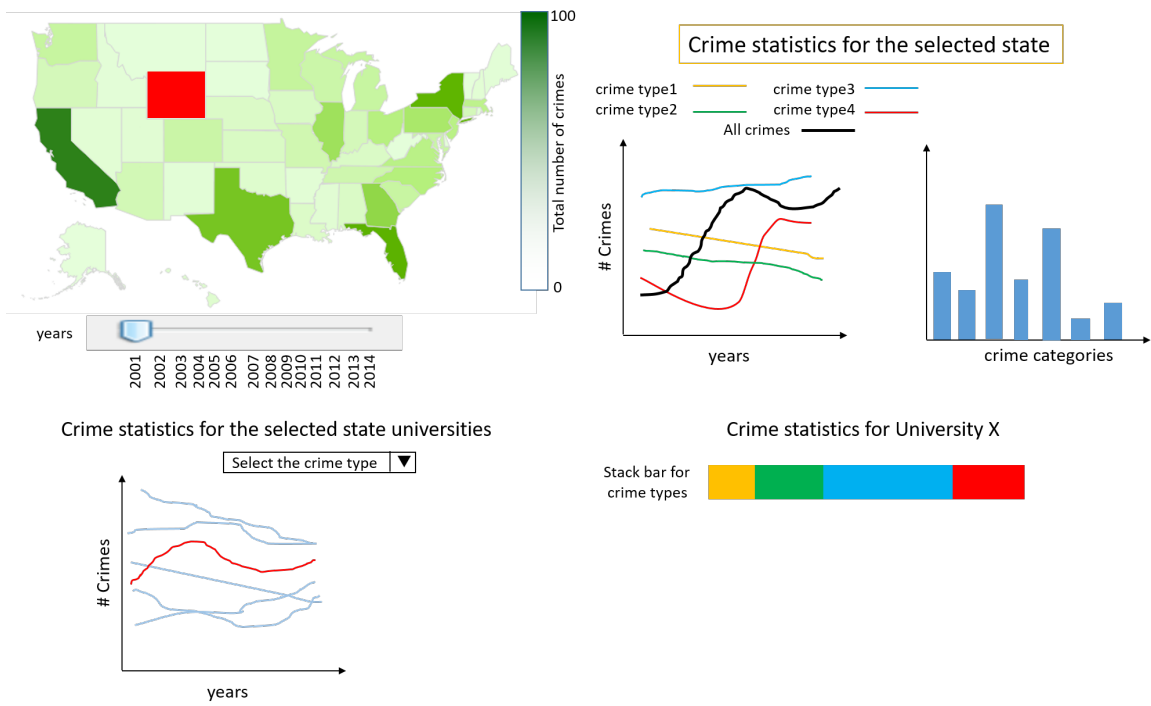
Figure 3:
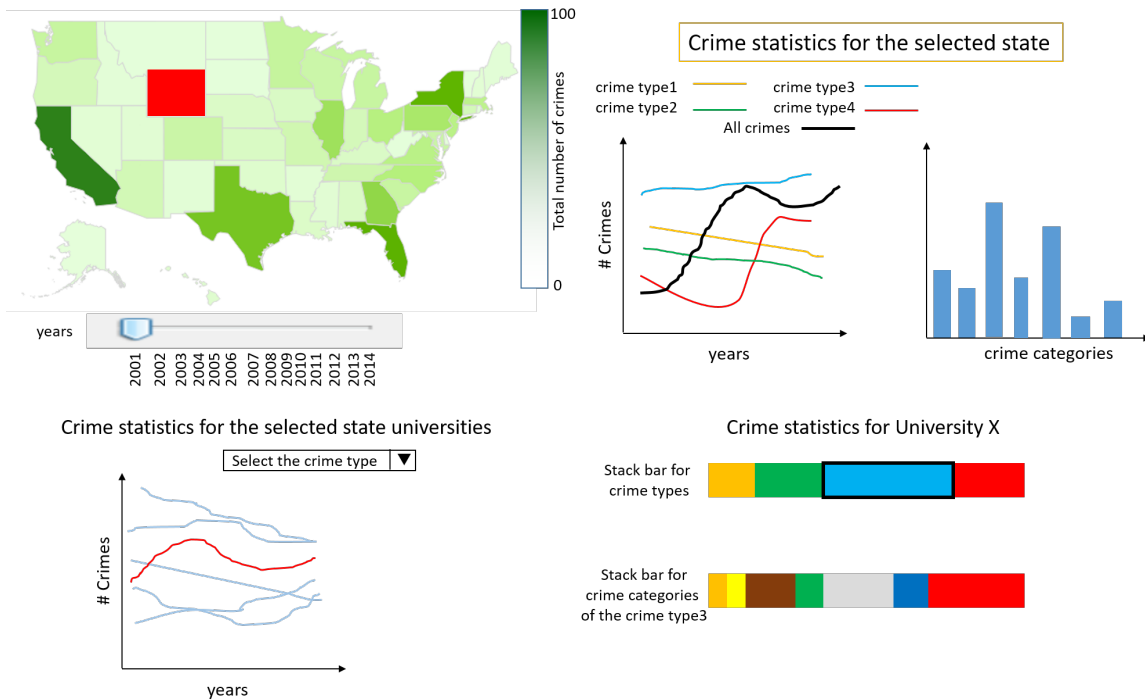

Figure 4:

Figure 5:

Figure 6:

Figure 7:

the x-axis represents the total number of crimes while the y-axis represents the number of students. Circles on the scatter plot represent the schools. Therefore, the ratio of the number of crimes to the number of students could be observed on the chart. These information will get updated as the slider for different years gets changed to see how these data get changed over the years. In this prototype in addition to selecting the states, you can use brushing on university circles, to update the selected university line chart and stack bar chart that we had in the first design as well (Figures 7 and 9).

4.4 Comparing different types of crime:

In this design instead of a map we display all the universities with circles on the screen. The size of the circles are proportional to the number of the students at each school (Fig. 10). They are colored using a gradient colormap that uses saturation to show the total number of the crimes at each school over the number of students. The location of the circle gives us a sense of the ratio of four different types crime compared to each other. Namely, if the circle leans on the corner representing criminal offense, the proportion of criminal offense in the corresponding school is higher. Here by hovering over a circle we can highlight it and see the name of te school bold and clear and also the exact number of crimes in that school Fig. 11. There would be a bruch selection to compare statistics of some desired schools in line charts (Fig. 12) and for a selected school on stack barcharts (Fig. 13).

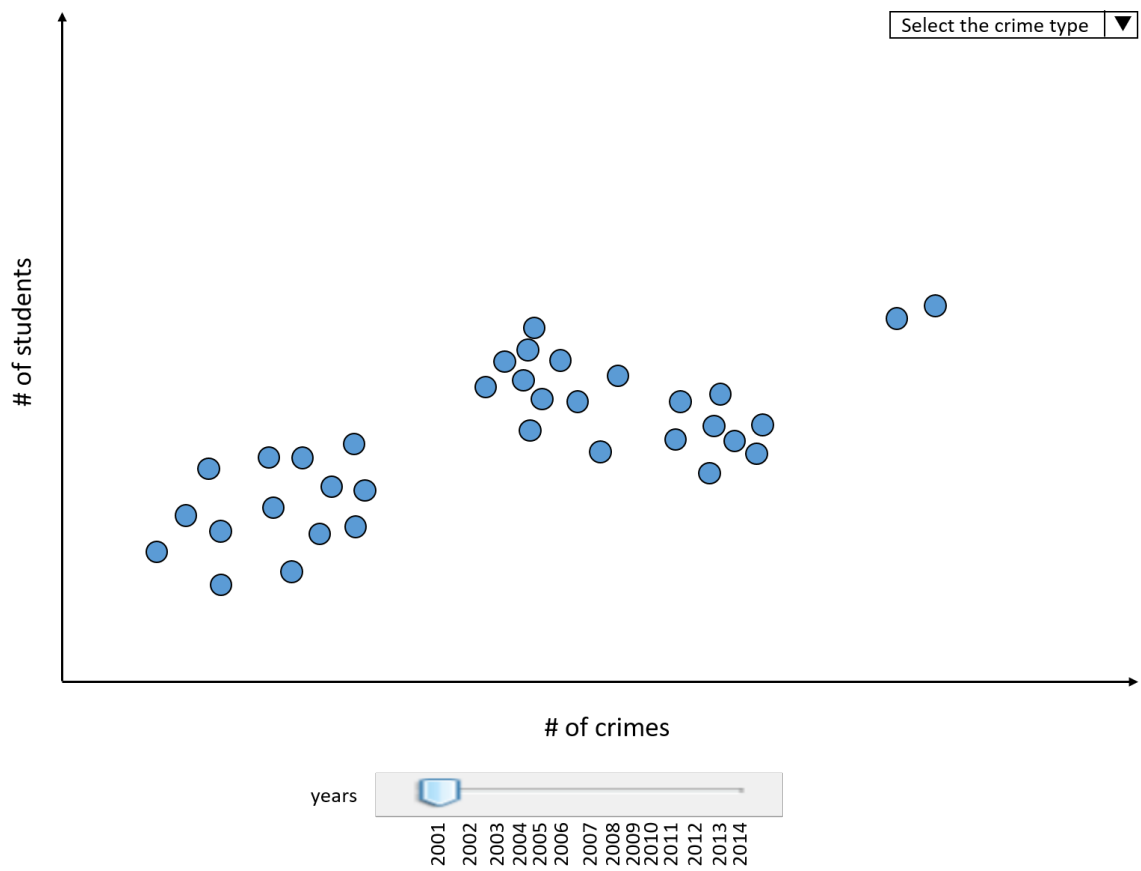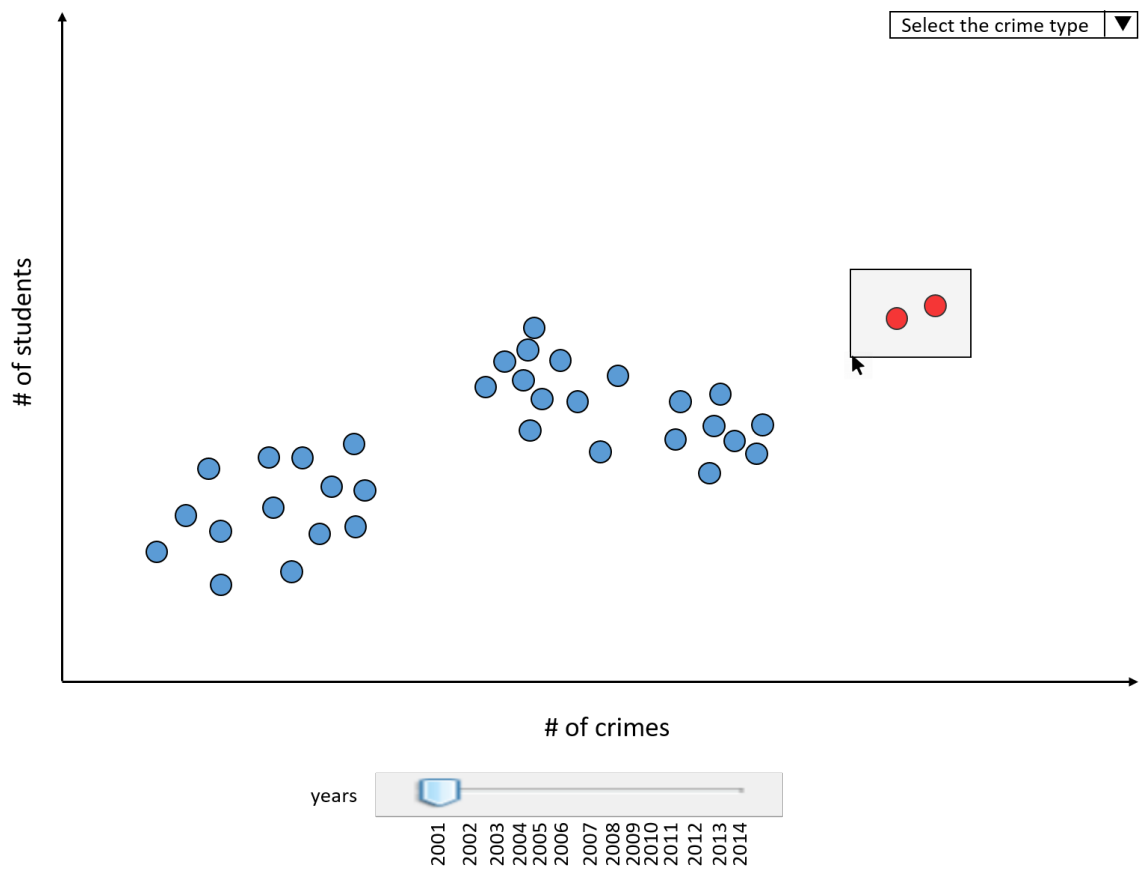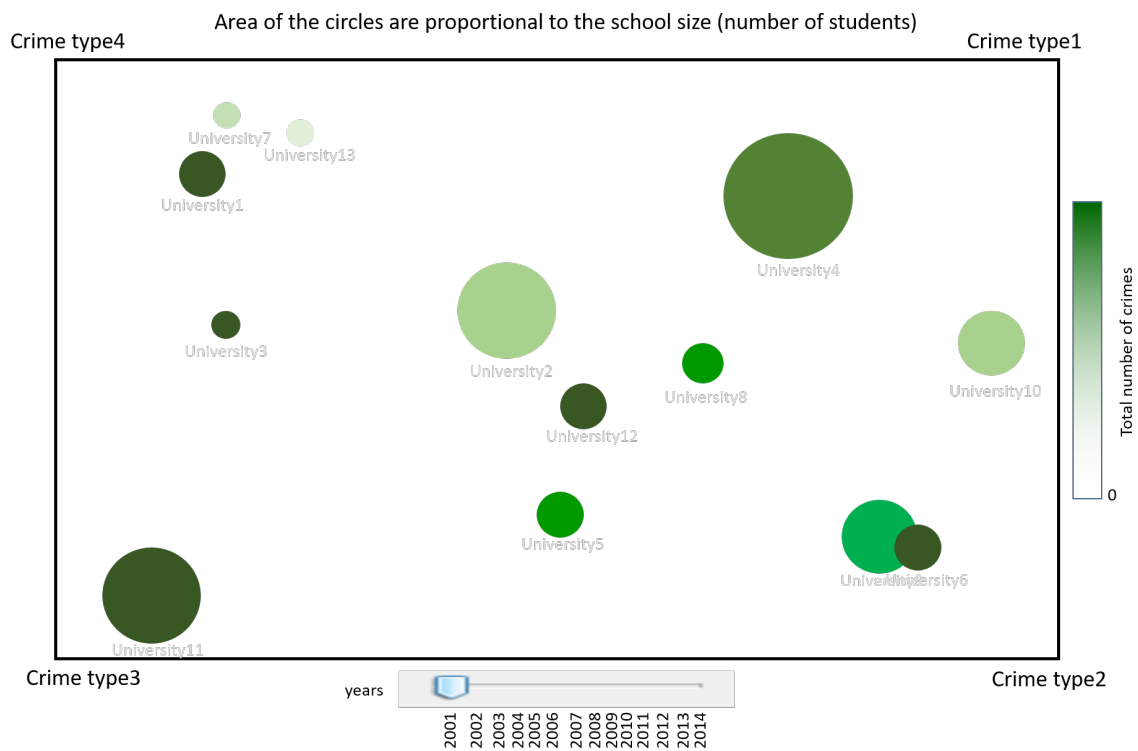The similarity map visualization presented in class inspired us to think of this design for this prototype. This visualization can be reached at "http://mariandoerk.de/edgemaps/demo/#music;map;;;".
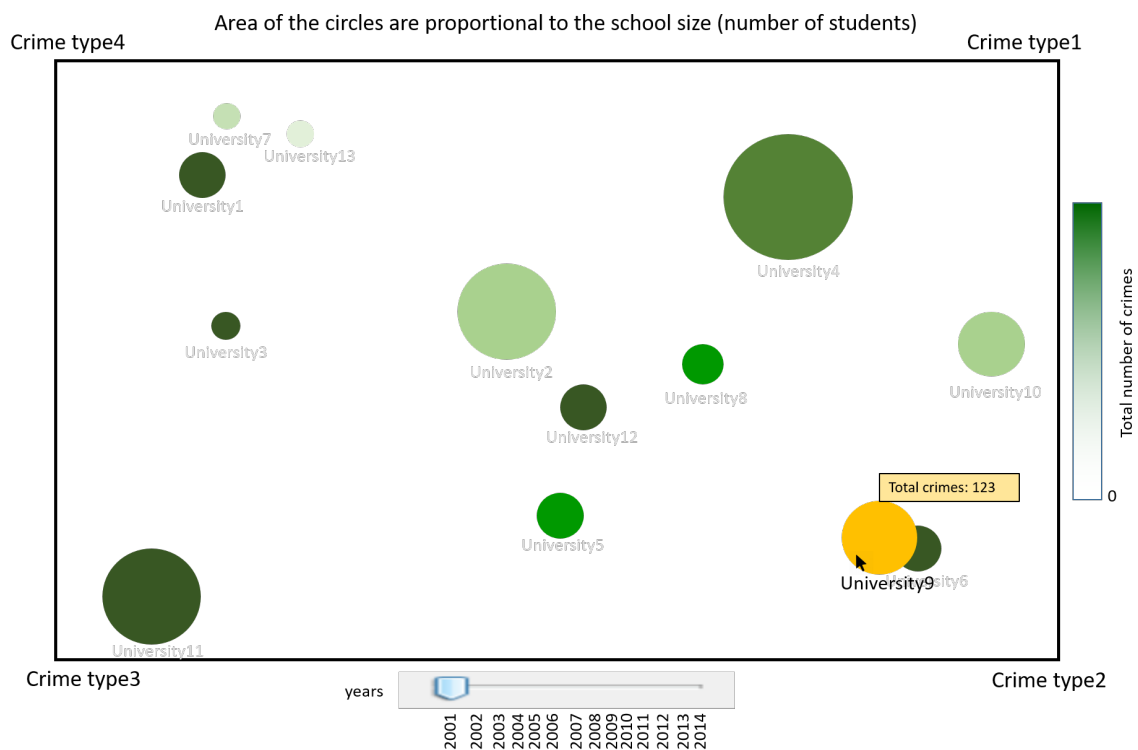
Figure 8:

Figure 9:

Figure 10:

Figure 11:

Area of the circles are proportional to the school size (number of students)

Crime type4

Crime type1

University7
University13
University1

University4

University3

University2

University8

University10

University12

Total number of crimes

0

University5

University9 University6

University11

Crime type3

years

2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014

Crime type2

Crime statistics for the selected state universities

Select the crime type ▼

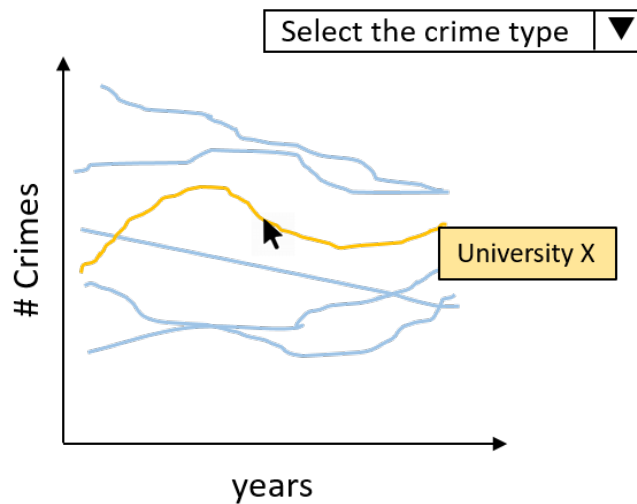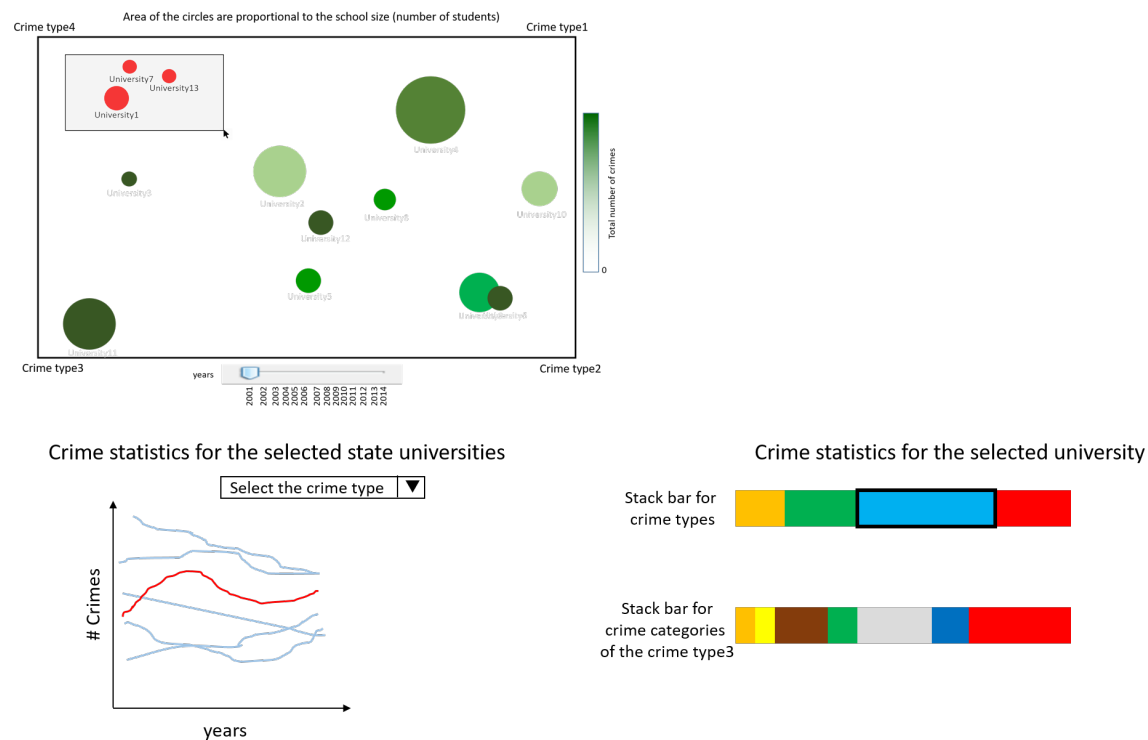# Crimes

University X

years

Figure 12:

Figure 13:

4.5 Our selected best design:

We chose the first design and the third design as our best designs. In both of these designs we use a gradient colormap to qualitatively show a continuous attribute (number of crime per number of students). Then we use position for the most important data (cumber of crime in different categories). We have also used bar charts for separating categories and showing their amount. The stack barcharts are also used to show the percentage of each crime type or category, that uses length, the second best options. These designs, use color (second best choice for categorical attributes) to separates different categories since the spatial region is not available. The third design also uses spatial region to show crime types.

But there is only one problem with the third design (Fig. 10), which is there is a probability that the circles of two universities lay on top of each other and make it impossible to separate them. Because of this issue we decided to go with the first prototype (Fig. 7).

4.6 Issues and critiques

According to the feedback and critiques we received from our peers and our assigned TA (Yogesh Mishra) there were some issues with our initial designs. Since there may be a lot of universities in some states like California, the line plot for showing them in Fig. 4 may not be a good option to do that. Therefore, we decided to use our third prototype described in §4.4. But we use it for a selected state here instead of all the universities in US. This will make the universities more distinct.

There was also a suggestion to combine our stack bar charts into one layered donut chart. This can be inspired by the visualization shown at "https://www.jasondavies.com/coffee-wheel/". In this way we can see the ratio of all the categories at once. We will also get better transitions when selecting one category. There was also a suggestion to combine our stack bar charts into one layered donut chart. This can be inspired by the visualization shown at "https://www.jasondavies.com/coffee-wheel/". In this way we can see the ratio of all the categories at once. We will also get better transitions when selecting one category. Figure 14 shows a snapshot of this visualization. The real visualization is interactive and suitable for categorical datasets.
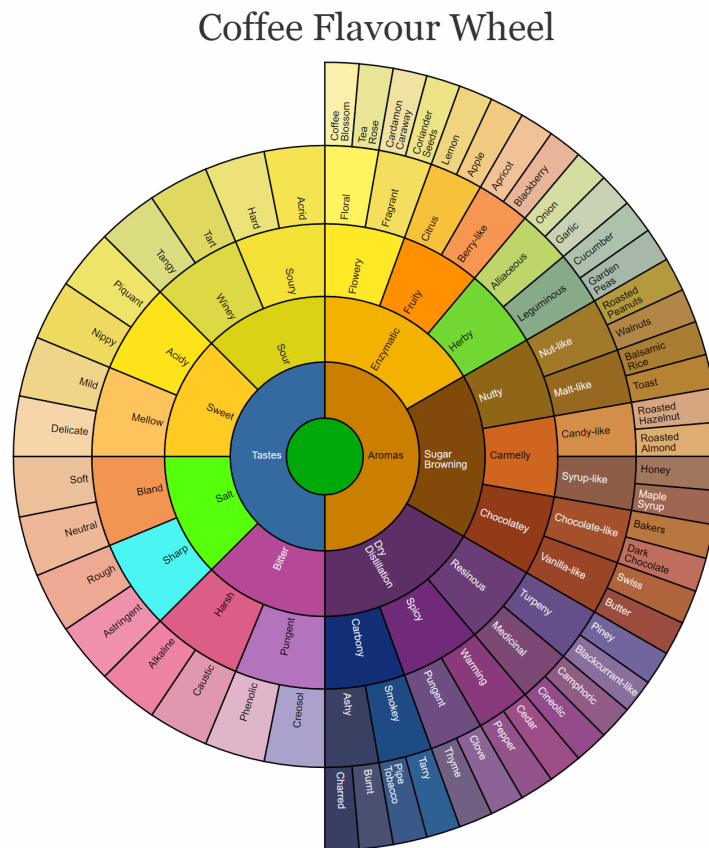

Figure 14: A sample of sunburst tree used as inspiration for showing crime types and categories.

We also decided to include a line for All of the crimes combined together on the line chart beside the map as you can see in Fig. 1.

Another issue that we thought of was for comparing universities. Previously we wanted to do that using two search boxes. But, only two search boxes for a large number of universities and comparing only two of them is so limited. Therefore, we decided to compare several universities in a table. For selection of the universities, we used a brush selection in the scatter plot of the universities similar to the plot in ??. In this way we can choose several universities simultaneously for comparison. We can also see the crime rate of the selected universities in more detail.

5 Implementation

The first view of this visualization is the map of United states. We can interact with the map and update other charts from here. It was crucial to implement this part correctly and make it ready for smooth transitions and synchronizations. For this part we used a topo-jason file to draw the US map which included state ID's. We passed the state abbreviations to drawn states on the map to have access to them later. We also attached a tool-tip which pops up when the mouse hovers over a state showing general information about it. A range input is implemented to handle the time slider below the map. This allows us to move across different years and access the related data. There is also an "onclick" event for each state. The user can click a state to select it and see its corresponding data visualized on rest of the charts. It all could be seen in Fig. 15.
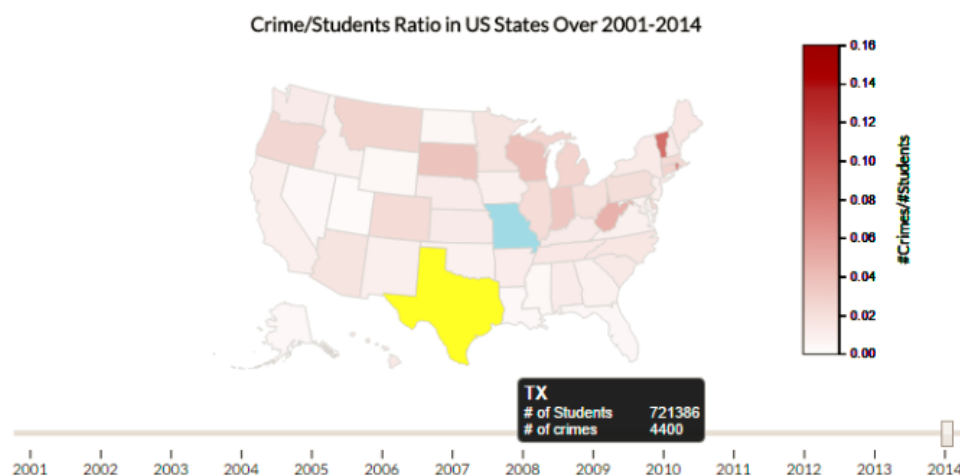

Figure 15:

By choosing a state on the map, two new charts would get updated. The first one is a line chart in which each line displays the stats for one type of crime in the selected state during all the years the survey has been conducted. By clicking on each one of the lines, a bar chart will appear showing the number of crimes committed in each category included in the selected type of crime. Dues to lack of space, the name of each category is displayed in a tooltip whenever mouse hovers over one of the bars. An example is shown in Fig. 16.
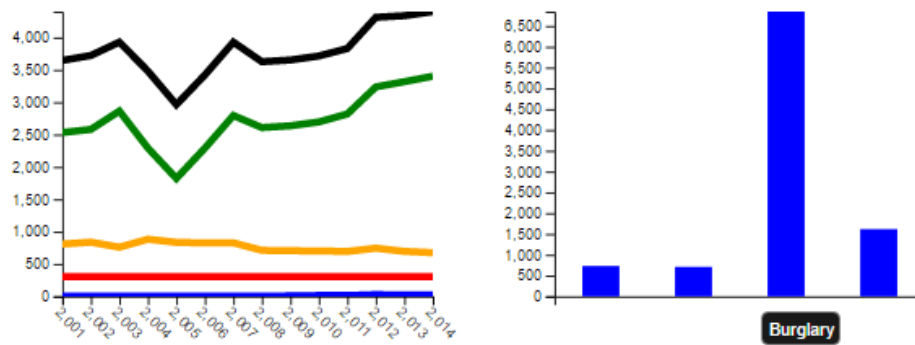
Figure 16:

The second chart which get updated after selecting s state is the ratio Chart. The institutes in the selected state would be displayed on a rectangular chart, each being represented by a circle. In this chart, each corner represents one of the four types of crimes. And as a crime type has a larger role in the total number of crimes in an institute, the said institute would be placed closer to the corresponding corner. In this chart the size of each university is displayed via the size of its circle. Also, color saturation is used to display the ratio of number of crimes in each institute to its number of students. An example could be seen in Fig. 17.
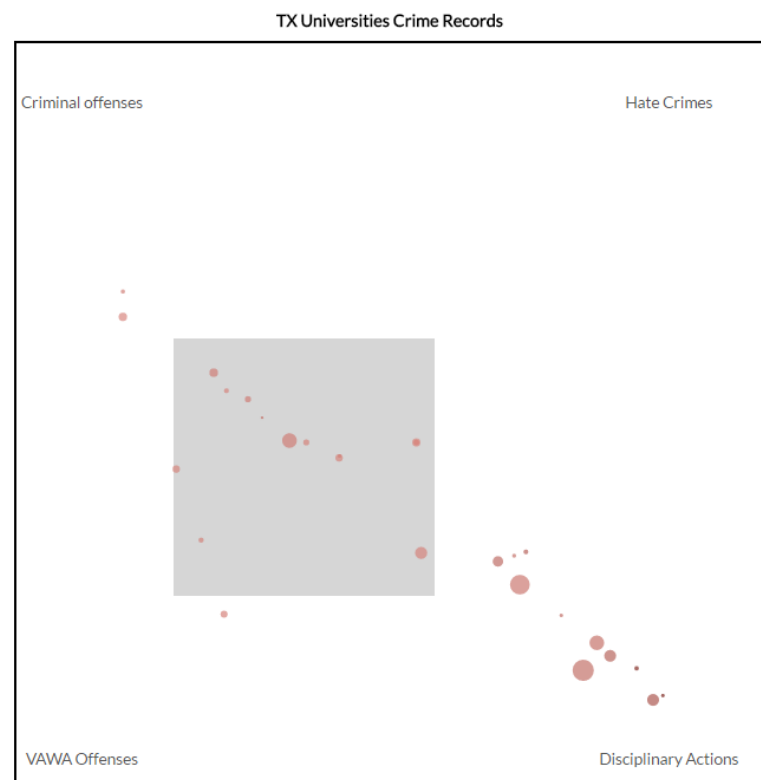

Figure 17:

By clicking on one of the circles on the ratio chart, the SunBurstTree chart would pop up displaying the full crime stats for the university which has been selected. This chart looks like a layered donut in which, the first layer represents the crime types and the second one represents the categories in each type. Also, by clicking on a a specific type of crime, the chart gets updated and as shown in Fig. 19 and by clicking on the inner circle it goes back to its initial shape.
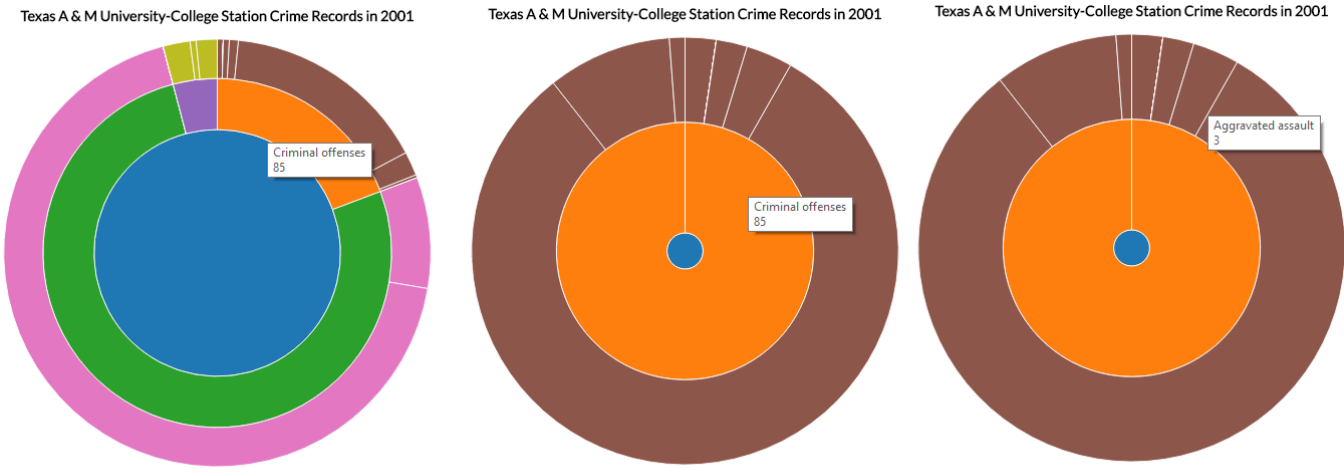

Figure 18:

By selecting a set of universities on the ratio chart via brush, the comparison table gets updated. As the name suggest, the goal of this table is to compare crime statistics in different schools Fig. 19:

Institution Name	Institution Size	Criminal Offenses	Disciplinary Actions	Hate Crimes	VAWA Offenses
Columbia College	12954	0	0	0	0
Park University	13259	0	0	0	0
Webster University	19398	0	0	0	0
Baylor University	14174	16	10	0	6
Texas A & M University-Commerce	8813	31	24	0	10
University of Houston-Downtown	11793	10	3	0	4
University of Houston	34663	78	44	0	33
The University of Texas-Pan American	17435	10	2	0	10
The University of Texas at Brownsville	17215	9	7	0	3
Texas A & M University-Kingsville	6567	32	14	0	13
The University of Texas at Dallas	14556	35	13	0	13
The University of Texas at El Paso	20154	52	13	0	19
The University of Texas at San Antonio	28533	19	31	0	11
Texas Woman's University	12168	10	4	0	16

Figure 19:

6   Evaluation

By working on this dataset, we found that the data records for the crimes in US universities is sparse and their data collection has not been done all years for all universities. As we go back later in time the amount of data reduces, which is expected.

Using this visualization we could answer all of our question regarding the safety in US universities. For example in 2014 Utah, Wyoming and Mississippi has been among the safest states with lowest crime to school size ratio, Utah has stayed among the safest states for all these years from 2001-2014. We also found that the Vermont was always the unsafest state. But we only had one big campus in Vermont which had about 128000 students and 1081 crimes in 2014. But there are other Universities in other states with higher crime ratio. The state ratio may get small because of some universities with low crime rate. For example West Virginia University had about 0.6 crime to number of student which is so high.

For someone trying to find a safe school in California, it is useful to know that the University of California-San Diego had the highest number of crimes in 2014. This is while Carnegie Mellon did not have any recorded crime for that year .

By looking at the scatter plot of the schools in a few states over the years, we will find out that the most of the crimes are related to disciplinary actions or criminal offenses. Universities with high hate crime numbers or violation against women are rare, which is fortunate. But there are still some outliers like several universities in Florida.

It can also slightly interpret from the scatter plots that the number of violence against women has decreased in time.

The other thing we can find is that what kind of crime a specific school has the most. For example, most of the crime in University of Utah is related to disciplinary actions, which are mostly drug and liquor related violations.

We can also see the trend of crime at different states. As instance, in Florida criminal offenses number has decreased in time while disciplinary actions number has increased from 2001-2014. Generally, the number of crimes has increased in Florida during this period. The number of crime in almost all the states has been increasing, but they have also got bigger in size, so slight increases should not worry us. However, there has been a period in around 2007 that Nevada has experienced a good reduction in the number of campus crimes.

This visualization can also be useful for social researchers, who are looking for the roots of the problems related to campus crimes in different geographical locations.

In terms of performance, our visualization runs smoothly at each step. We also tried our best to follow the principles we have learned during the course to choose the best tools available at each level. Using our implementation, it is easy to look for crime statistics in different schools across United States in a variety of ways and also compare different states and school in terms of their crime rates. However, we believe there are two parts missing that could improve our results. The first one is the ability to filter the institutions based on the user's needs. We skipped this part due to the sparsity in the data set specially in the smaller sized institutions. The second part is the ability to compare any two different schools across the US. For now, it is only possible to compare the schools located in the same state. This part was skipped simply due to lack of time.