

PRESS BOOK

(Linghong Chen, Suhas Watturkar, Peter Gathua)

I. OVERVIEW

(April 3, May 5, 2015, Suhas, Linghong)

Background

Campus safety is of utmost importance to everyone. There are a number of sources from where safety data is available, but these sources do not necessarily provide an integral view of the data, which can be easily correlated. Also, the magnitude of this data is large and the granularity provided by the data is also very diverse, which makes it very difficult (if not impossible) to make informative conclusions from the data.

One prominent sources of campus safety data is The Office of Postsecondary Education, they provide this data at granular level, with different severity levels right from disciplinary actions to criminal offences, on / off campus locations, etc. This data is available for years 2006 to 2013. The goal of this project is to unify this data into a common data semantics, and provide interactive visualizations for this data.

Motivation and Goals

The objectives of this project are to provide visual representation of the campus safety and security data in following two major categories:

a) Map View

The user will be able to interact with the visualization and review the data based on various dimensions, such as location (state, county, city, etc.) or individual college.

b) Crime patterns view

The crime pattern view will enable the user to visualize the possible correlations with college details such as ranking, acceptance rates, gender ratio, size, etc.

Related Work

A century of Corn (<http://jbencook.github.io/cs171-ag-viz/src/CountyData.html>)

Despite we focus on different items. But our site and their site both want to discover nationwide distribution of a target item and historical data using a map, and we both also want to dig correlations between our target item and the external factors.

Questions to Answer

The visualization should answer the following questions:

- a) Which area/state schools have more severe crime fact
- b) Which individual schools have more severe crime fact
- c) Does one school category (such as two-year colleges) had more violation than another category?
- d) Does other factors affect the college campus security(such as the gender ratio, school size, college quality (admission rate and ranking)?

II. DATA SOURCE

(March 27-April 10, 2015, Linghong, Suhas)

Campus safety and security data is available from the Office of Postsecondary Education (<http://ope.ed.gov/security/>) in excel format. From this data source, following two data set files were used:

File name

oncampusarrest

Description

Contains on-campus arrest data for various categories of crime ranging from drugs to weapons related offenses.

oncampuscrime

Contains on-campus crime data for various categories such as murder, robbery, sex offense, burglary, etc.

Each data set covers data from all the United States colleges for a three year period, starting from the earliest (2006-2007) to most recent (2011-2013) periods.

Each dataset has the following common portions: school name, address, and number of students by gender, total student number, school category (public, private, 2 year, 4 year, etc.).

Each dataset also contains security information. The **on-campus arrest** data files have weapons, drugs, and liquor crime categories, while the **on-campus crime** file has nine crime categories.

In our initial plan, we also planned to use other data sets, such as the file sets for **oncampusdiscipline** which contains discipline data for the categories under on campus arrest dataset, and a .sav file for the college application data from Open ICPSR that includes college name, number of college applicants (male and female), acceptance rate, the college's US NEWS college ranking, and the college's student body size. However, due to various reason, we changed to only focus on the **campusarrest** and **campuscrime** data set. (please see the details in the data analysis chapter.)

A US map geo code:

From Scott Murray"s book " Interactive Data Visualization for the Web",
<http://chimera.labs.oreilly.com/books/1230000000345/ch12.html>

US zip code geo code:

III. DATA CLEANING AND CONVERSION

Choosing the Right Dataset

(April11, 2015, Linghong)

The security and safety data has on campus, residence halls, non-campus and public property four groups of dataset. However, the security data for non-campus and public property have low number, which makes difficult to distinguish whether a change is a random variation or a real pattern. For example, weapon possess is one of the data set that has low crime number. From the four total weapon possess charts (fig1 a-d), we can see three of these charts have somewhat downward trend, but fig 1c has very low crime number and a high random variation, which makes difficult to discover any pattern or trend.

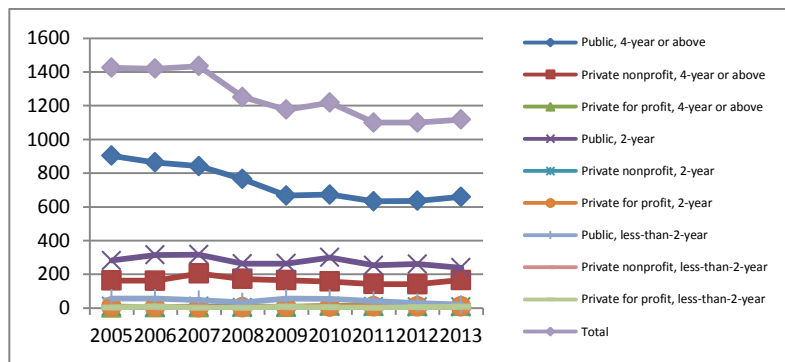


Fig1a total illegal weapon possess happened on campus.

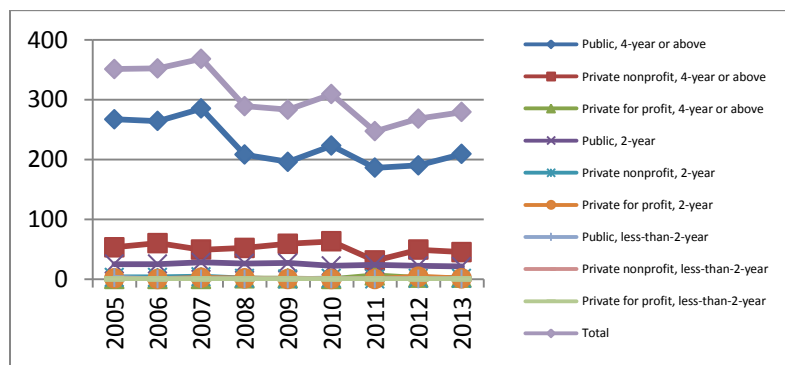


Fig1b total illegal weapon possess found in college residence halls

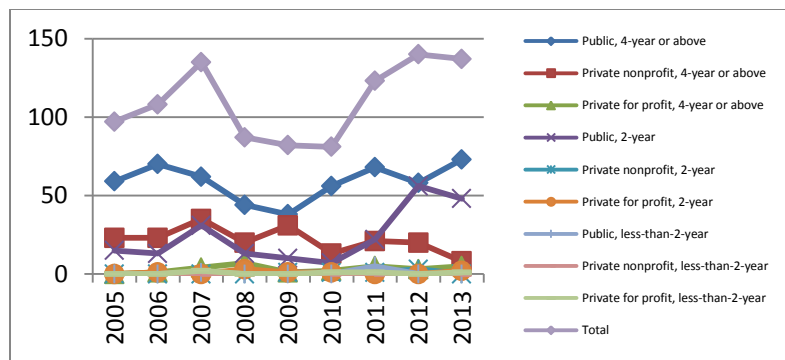


Fig1c total illegal weapon possess found in non campus area

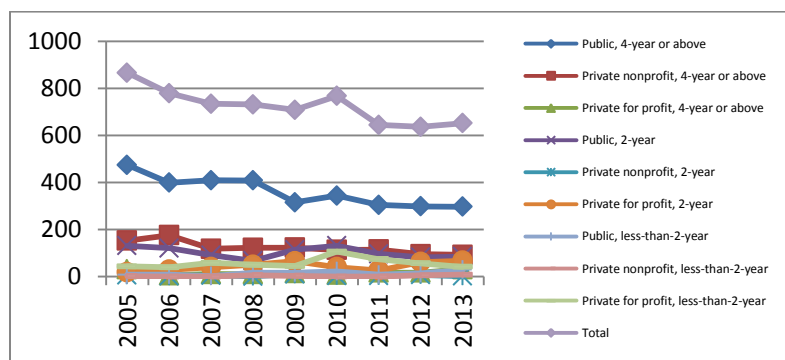


Fig1d total illegal weapon possess found in public property

The above charts also provides a clear indication that compared other groups of dataset, on campus data has relatively high number. Research further of the other data groups that has low numbers gave us similar conclusion. Also considering that in some school categories (such as community colleges), students usually don't live in campus resident halls, thus comparing the data in residence halls could cause bias, hence, we decide to only focus on the on campus security and safety data.

The on campus data has arrest, crime, discipline and hate four degree of violations. Arrest and discipline has three crime categories, while the other two have nine categories with arrest and crime categories have more severe situation than other two categories, thus for a project with limited time frame, we decide to make a site only focusing on campus arrest and crime.

Data Cleaning and Conversion

(April 11, 2015-April 16,2015, Linghong and Suhas)

We first combined on campus arrest and on campus crime data into one file. They were totally 12 categories. Two file covers three year period, the earliest one covers two year period. Then all schools whose state columns are empty were deleted. Those schools are the oversea schools established by US colleges.

Further analyzing the data, we found the data also includes colleges in Puerto Rico and US virgin islands. They have state name, hence, we didn't notice them in the beginning. Those schools thus were deleted from our data set too. About 10% of colleges didn't submit security data. Those schools

were also deleted from our list. The clean data were then sorted and linked to geo coding number (see below). The two recent ones use same University ID, were then combined them into one file in Access, and convert it to a JSON file. The 2006-2007 data has different way of naming college ID, we thus didn't combine it to the above file.

Connect Campus Security Data to Geo Data

(4.10-4.17: Peter)

We planned to combine college address to geo coding address, but we encountered some difficulty, thus three college crime data files ranging from 2006-2013 were linked to geo data in the zip code level using sql, and csv files were exported.

IV. DATA ANALYSIS

Data Analysis Find Interesting Patterns and Trends

(4.14-4.17, Linghong)

The initial try of analysis of the data didn't turn into precise result because we didn't notice there were schools belong to Puerto Rico and US Virgin Islands. We use the clean data reanalyze the data again. Since each category has very different number of schools and students, we use total school number or total students number in each school category to normalize the data. The data analysis provides a clear indications of following trends:

I. On Campus Arrest and Crime Number Is School Sector Related

The edu.gov site provides 9 years of security data. To save time, we only did an analysis of the most recent 6 years for 6 school sectors that are generally more interested by students, parents and teachers.

Research of the on campus arrest data showed a clear crime pattern related to school sectors(Fig2a-c). Generally speaking, public 4 year colleges have much higher number of on campus crime issue compared to other 5 college categories. Since the school number between each school category varies significantly, the data shown here are normalized by school number, and listed as crime case number per 100 schools in each school sector. The school sectors are: 1: Public, 4-year or above, 2: Private nonprofit, 4-year or above, 3: Private for-profit, 4-year or above, 4: Public, 2-year, 5: Private nonprofit, 2-year, 6: Private for-profit, 2-year.

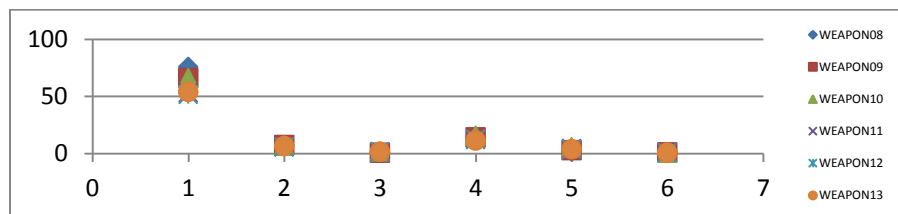


fig2a. On campus arrest for weapon violation during 2008-2013

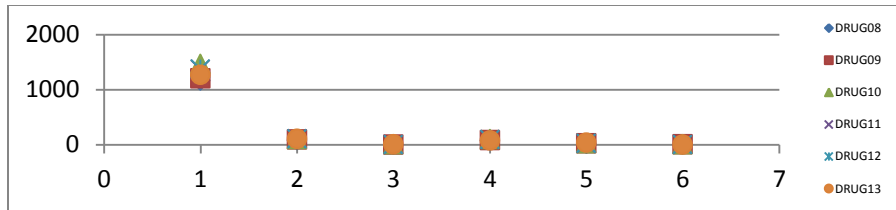


fig2b. On campus arrest for drug violation during 2008-2013.

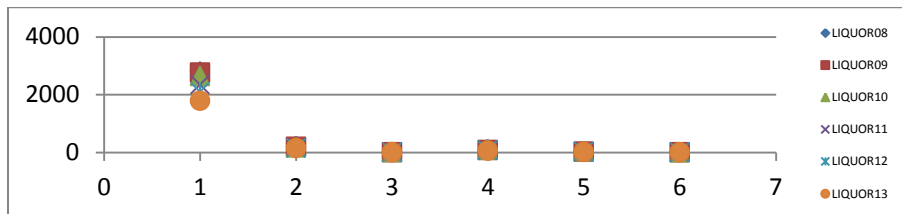


fig2c. On campus arrest for liquor violation during 2008-2013.

To analyze the nine categories of on campus crime data, we selected six categories that have relative large number for analysis. Interestingly, we also see a distinctive pattern that public 4 year schools have much higher crime number compared to other five college sectors(fig2a-f). Same as above, the listed number are shown by cases number per 100 schools in each school sector. The school sectors are as below: 1: Public, 4-year or above, 2: Private nonprofit, 4-year or above, 3: Private for-profit, 4-year or above, 4: Public, 2-year, 5: Private nonprofit, 2-year, 6: Private for-profit, 2-year

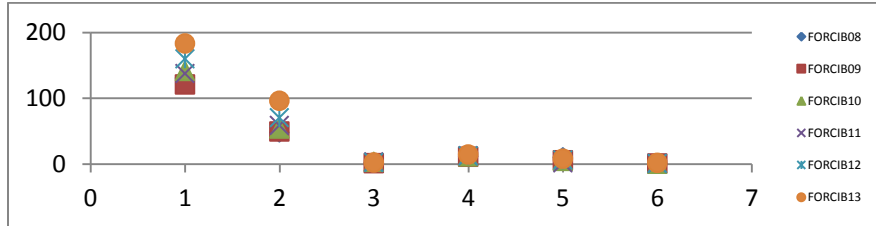


fig3a. On campus crime for forcible sex offense between 2008-2013.

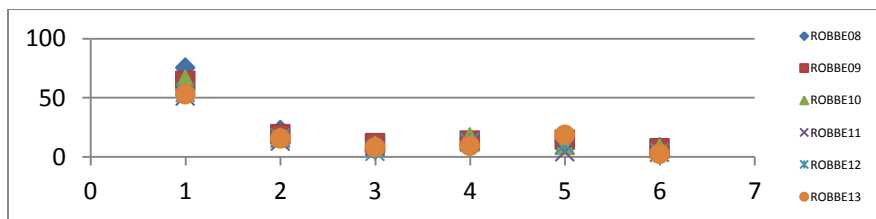


Fig3b On campus crime for robbery violation during 2008-2013.

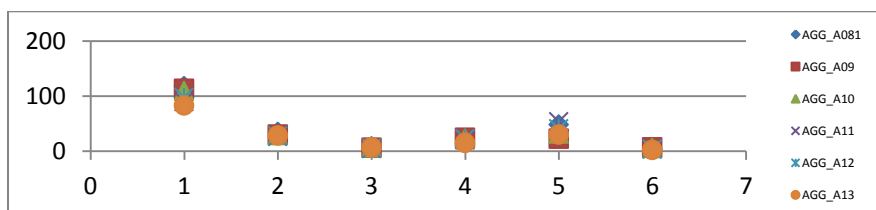


Fig3c On campus crime for Aggravated Assault during 2008-2013.

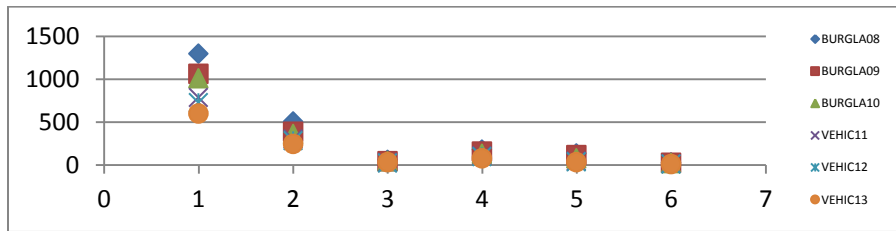


Fig3d On campus crime for burglary during 2008-2013.

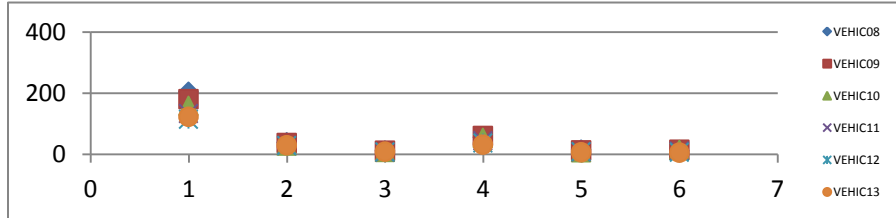


Fig3e On campus crime for vehicle theft during 2008-2013.

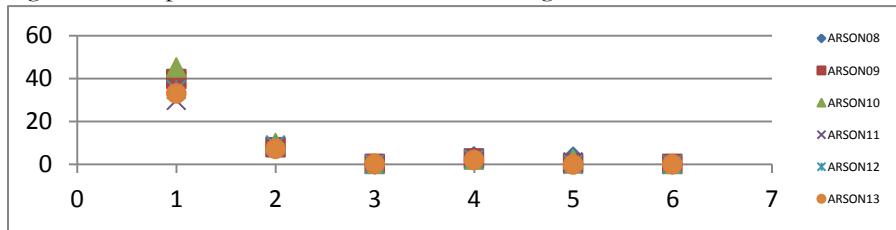


Fig3f On campus crime for arson violation, 2008-2013.

This above school sector related crime pattern is across all the nine on campus crime and arrest categories we have done the analysis, thus we believe our finding is real. We also assumed that the likely same pattern would exist in the state level. **We thus decide one of our box will be used to display the relationship between school sector and crime safety data.**

II. Yearly Trend: Public School Sector Gets Better, Other School Sectors Have No Change

Further analysis also displayed an interesting trend of public school crime data: among nine categories that we did analysis, six categories have a downward trend of crime number in the recent years. In contrast, there is no visible change seen in other school sectors.(Fig4a-i)

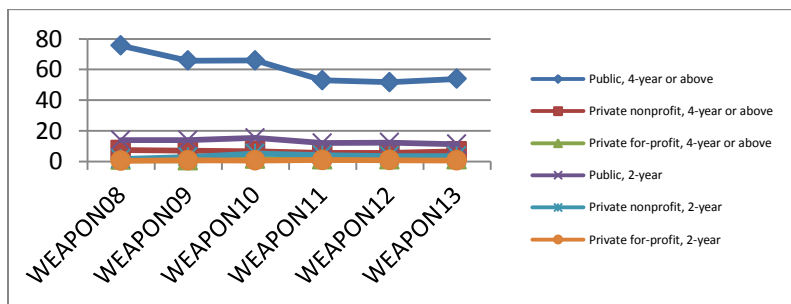


Fig4a On campus arrest for weapon violation during 2008-2013. There is a trend of decrease of crime number in public 4-year colleges.

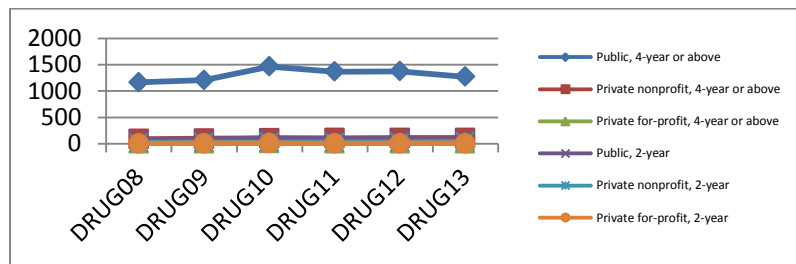


Fig 4b On campus drug violation during 2008-2013. Unlike the other category, public 4-year colleges have no improvement in drug violation.

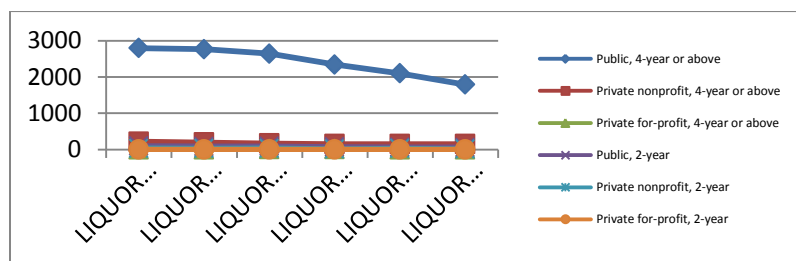


Fig4c On campus liquor violation during 2008-2013. Public four year colleges have a trend of decrease of violation number.

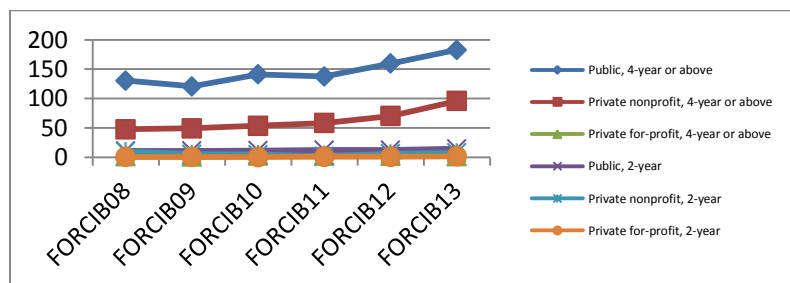


Fig4d On campus forcible sex offense data showed that public four year colleges have a trend of increase of violation number.

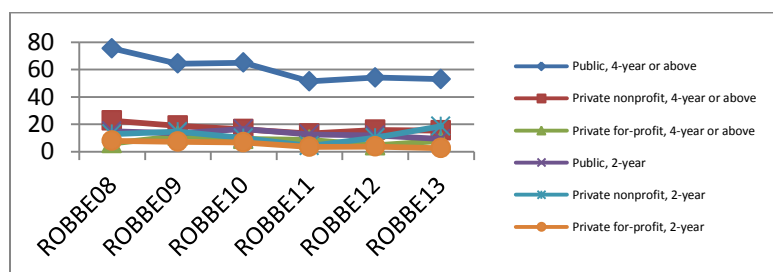


Fig4e On campus robbery violation during 2008-2013. Public four year colleges display a trend of decrease of violation number.

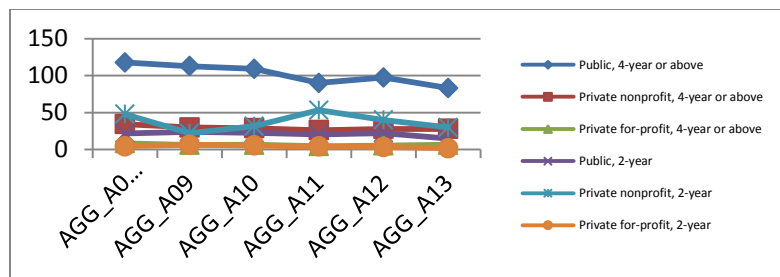


Fig4f On campus agg violation during 2008-2013. Public four year colleges have a trend of decrease of violation number.

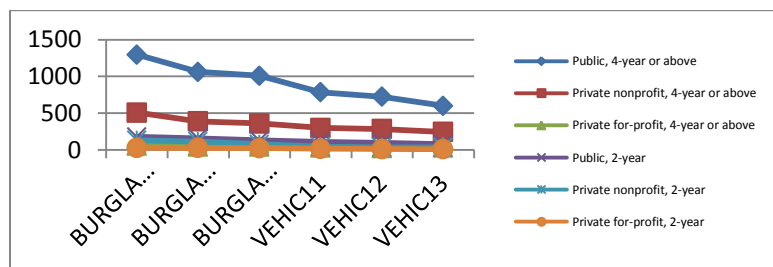


Fig4g On campus burglary during 2008-2013. Public four year colleges have a trend of decrease of crime number.

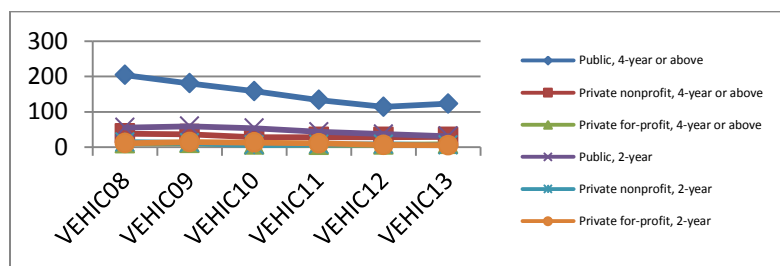


Fig4h On campus vehicle theft during 2008-2013. Public four year colleges have a trend of decrease of theft number.

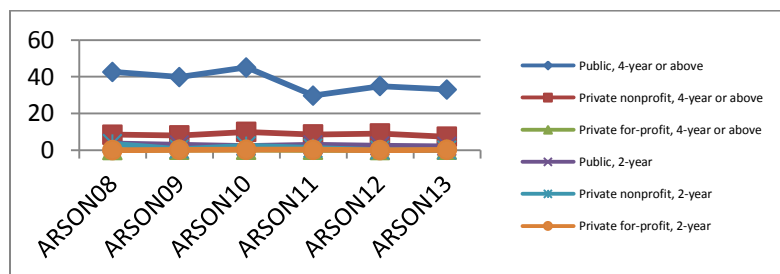


Fig4i On campus arson violation during 2008-2013. Public four year colleges have a trend of decrease of violation number.

Despite we didn't see all on campus security categories have downward trend in public 4 year college sector, we still believe it is an interesting finding, and a nice compensation to the finding that public 4

year colleges have high per campus crime data. **Thus , if time is granted, we may implement the yearly crime trend and display in one of our boxes.**

III. Possible Student Number Related School Sector Pattern

We also did analysis by normalizing the security data with total students in that college category. The results also displayed a distinctive pattern. In short, the three campus arrest groups have a pattern like fig 5a, while various campus crime data have a pattern like fig 5b. More importantly, for all of these patterns, we see a consistent high crime number in category 5 across various crime categories. (1: Public, 4-year or above, 2: Private nonprofit, 4-year or above, 3: Private for-profit, 4-year or above, 4: Public, 2-year, 5: Private nonprofit, 2-year, 6: Private for-profit, 2-year)

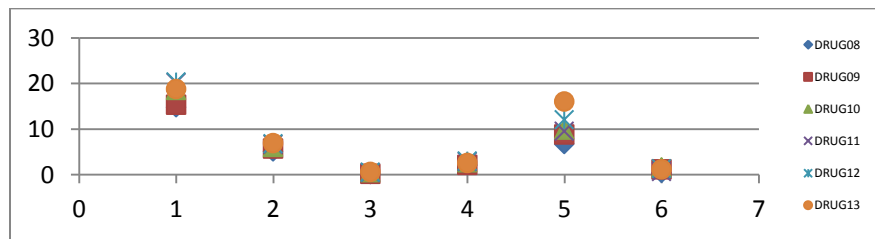


fig5a. on campus drug violation happened during year 2008-2013. Data are listed by crime cases per 10000 students

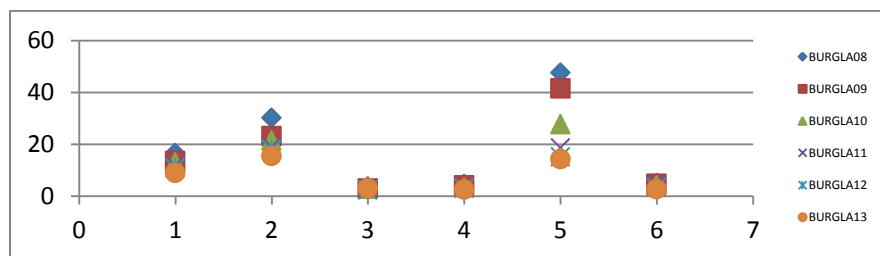


Fig5b on campus burglary during 2008-2013. Data are shown crime cases per 10000 students

One thing we have to mention is that the student number, female and male number listed in the dataset is not for each university campus, instead it is for each university, however, the security data provided is for each campus. Since not all university campus provides security data, the way of normalization by total student number could produce imprecise result. College sectors with more universities campus that didn't submit security data could benefit of this normalization and display artificially lower data. Thus, we think the campus security data do show some possibility of school size related school sector pattern, but have to be implicated with cautious.

Because of that, we won't implement this finding into our site. For the same reason , we also won't research further for the possible relation between the school size/female male ratio and the campus security in our site.

V. EVOLUTION OF THE VIS DESIGN

The First Vis Design

(3.31-4.3 Linghong)

Our first Visualization design were implemented a structure of two portions – Security Map and Visual Analysis of the data. (Fig6a)

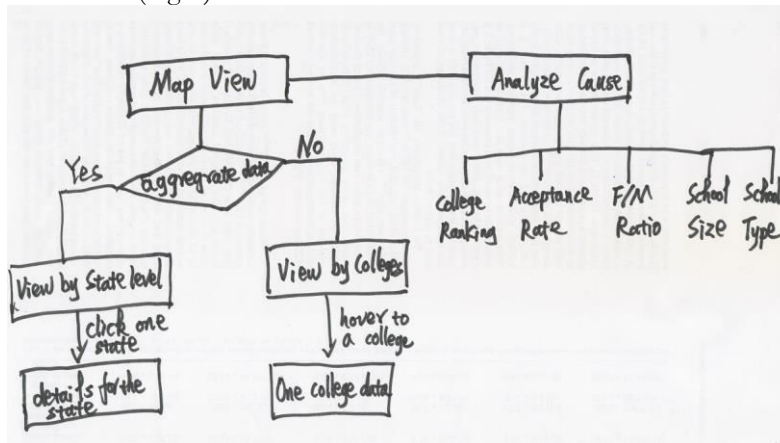


Fig6a

1. Security Map

The map was planned to show college security data to a state level, township level and college level. The former two will be area view and the last will be point view.

On the top of the map, a bar plot were planned to show aggregated data for various categories, which relate to the overall data shown on the map. Hovering on the individual bars will update the map with the data corresponding to the category shown by the bar. Similarly, the other radio buttons on the top side of the map will update the map with the category represented by the radio button. Hovering over the circle will show the number. We may also use other criteria to select colleges to show the security level (Fig6b, Fig6c)

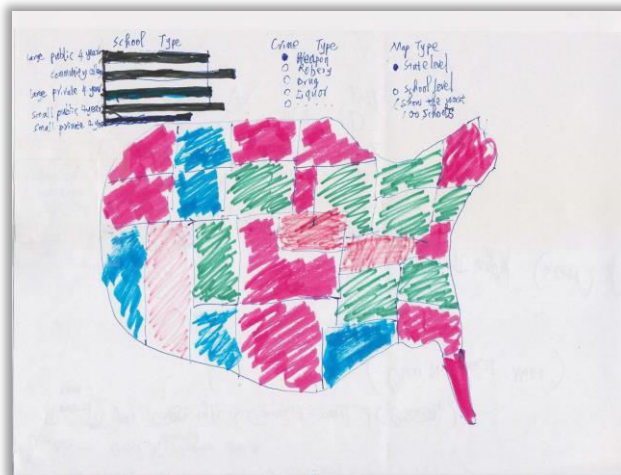


Fig.6 b Map Area view



Fig.6 c Map Point View

Clicking on a state will let the user *drill down* to the state level, showing all colleges within the state either using the zooming method (If we are able to achieve it) or use scatter plot (x as longitude and y as latitude)

2. Pattern Analysis

The visualization planned to discover the following questions:

- Does college quality (admission rate and ranking) affects campus security?
- Does one school category (such as two-year colleges) had more violation than another category?
- Does the gender ratio affect campus security?
- Does school size affect campus security?

Possible approach 1: Scatter plots may be used. Data will be grouped by several groups by different comparing factors. Within each group, random sampling will be used to reduce samples crowd level (see a example of graph below figure6d)

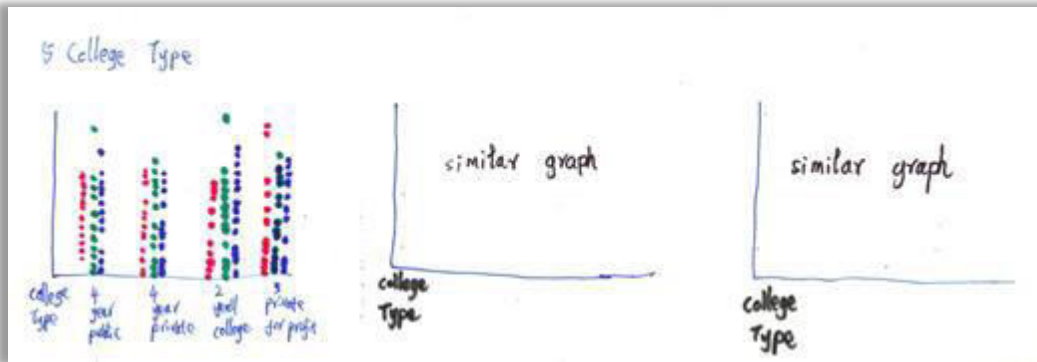


Fig. 6d a scatter plot analysis

Possible approach 2. Parallel Coordinates to show the relationship (Optional)

We will try the parallel coordinate method to see whether a clearer relationship among the above factors is found. (fig 6e)

(See an example at: <http://blocks.org/jasondavies/1341281>)

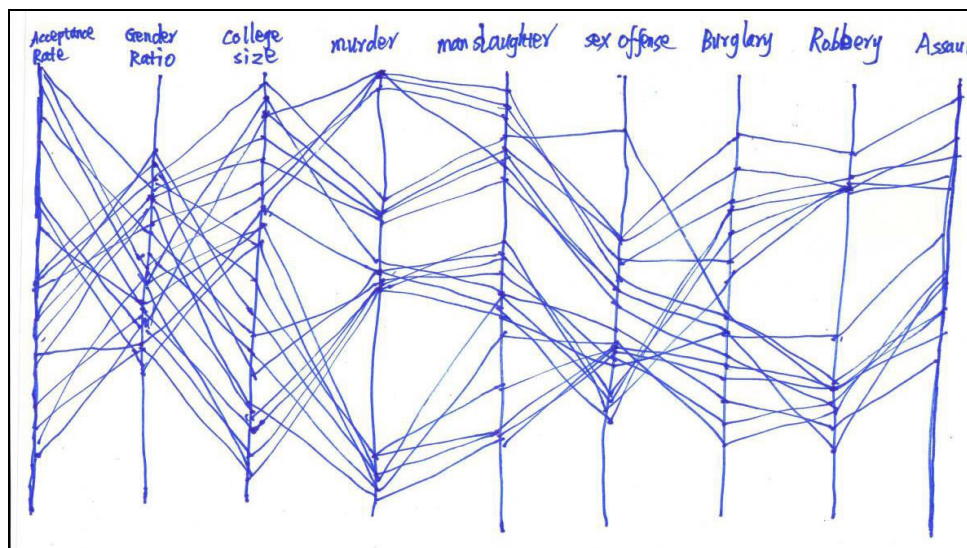


Fig.6eParrell Cordinates

The Second Vis Design

(April 19-21, Linghong)

Two things were going on: we get some distinct college crime trends, we needed to think about how to implement the trends. Also a new area that is for users to engage in was created. This area let users add their own judgment to give different weight on each crime categories, and the system then compute the total crime score.

According to these two facts, the site structure were planned to change in the following way: the map is still in the top center of the area. On the top right side column is a control panel. The rest areas will have three boxes, used for: 1) crime distribution across the 50 states, a multi-series scatter plots to show the school type related crime pattern and a multi-series line charts to show year trends. The map and all these three charts have to be interactive. Clicking one button trigger change in all four vis. (Fig 7a)

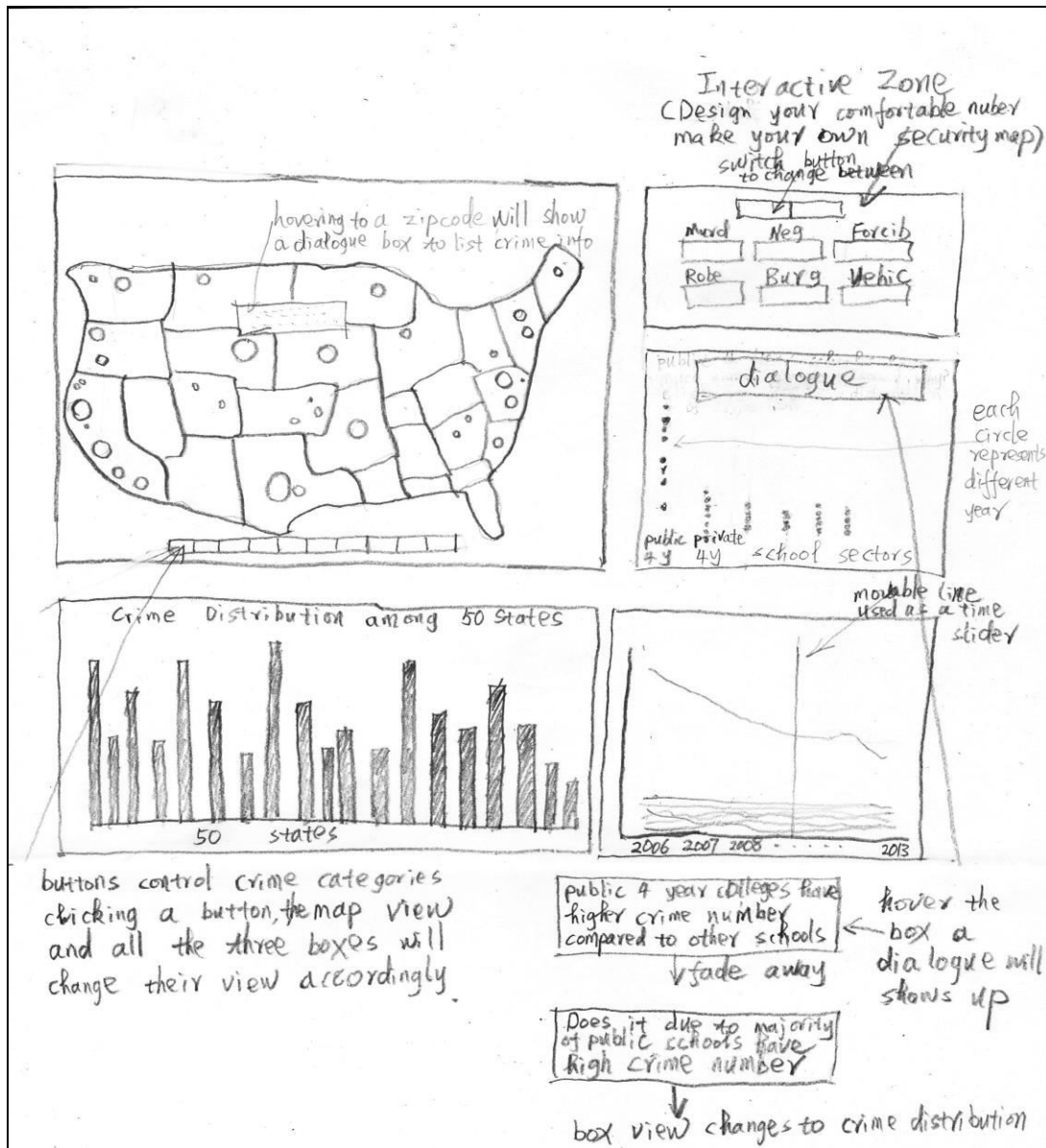


Fig7a

Besides, hovering on the scatter plots will also show a legend box to explain what the scatter plots charts mean. The legend should encourage users to click to find the answer why public 4 year colleges have much higher crime rate compared to all other college types.

Considering majority of colleges actually have very low crime rate, one explanation to the high crime rate seen in public four year colleges could be due to the fact that there are a few of public 4 year colleges have very higher crime rate, while in other college categories, those number is lower. For such purpose, a distribution bar charts could be used (Fig 7b). However, considering we don't have much time left, we are not sure whether we will have time to dig it so deeply, but let's first keep that plan in our mind.

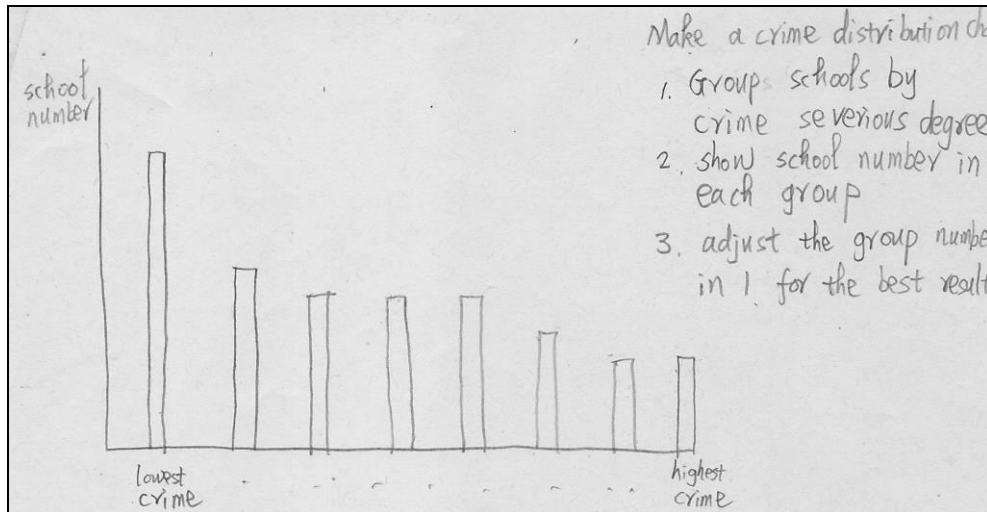


Fig 7b

The Third Vis Design

(April 28, 2015)

Our map will be used for two different purposes: showing crime trend and pattern, vs. let users to engage in making their own security map. We feel put into two different purposes of the work into one page causes confusing, thus we moved the map for purpose 2 to another page.

Front page

Inner Page

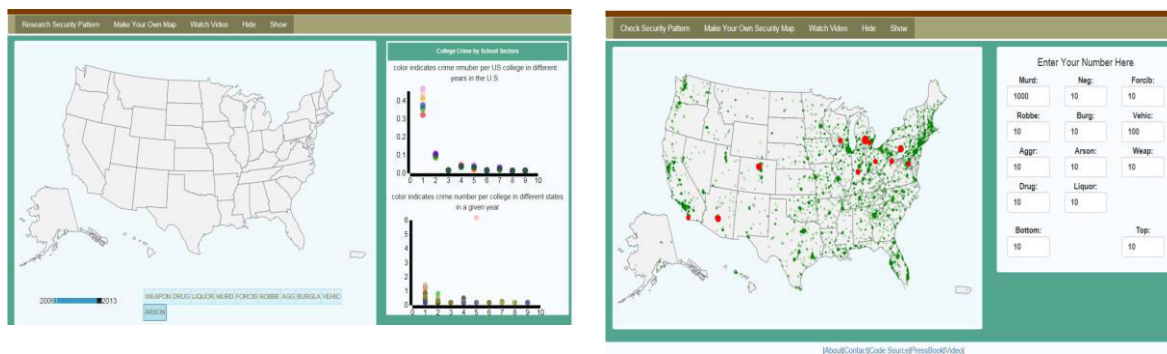


Fig 8

The Fourth Vis Design

(Linghong, Peter, Suhas, May4th, 2015)

Listening to our TA, Benjy 's suggestion, we consolidated our two pages' site view back into one page. (Fig9)

The map functions two purposes:

1) **Crime pattern map view with users' involvement** (The control panel is on the up-right area)

Users can input their judgment about which crime categories are more important to them and give higher weights on these crime categories. The system then calculates a total score for each colleges. This total score will be used to rank the colleges. The system then use the ranking data to calculate the circle size in the map.

2) Map view without users' involvement

The system can also show a single crime category without users' input. The volume is show the circle size. Because we found some colleges reported their schools in different zip codes but close location with same zip code. Also over the time, a zip code for a location could be changed, thus we didn't sum up the data for the same zip code. If the zip code represents the exact location of a school, it is better to show the crime volume in a zip code level, however, in our case, when more than one colleges are located in the same zip code, we made the color of the circle to be darker. This is not a perfect approach, but in our case, we have no other choice.

For the Map purpose 2, clicking a choice button beneath the map will trigger the change of map and other charts' view; For the Map purpose 1, adjusting control panel only changes the map view and the parallel coordinates.

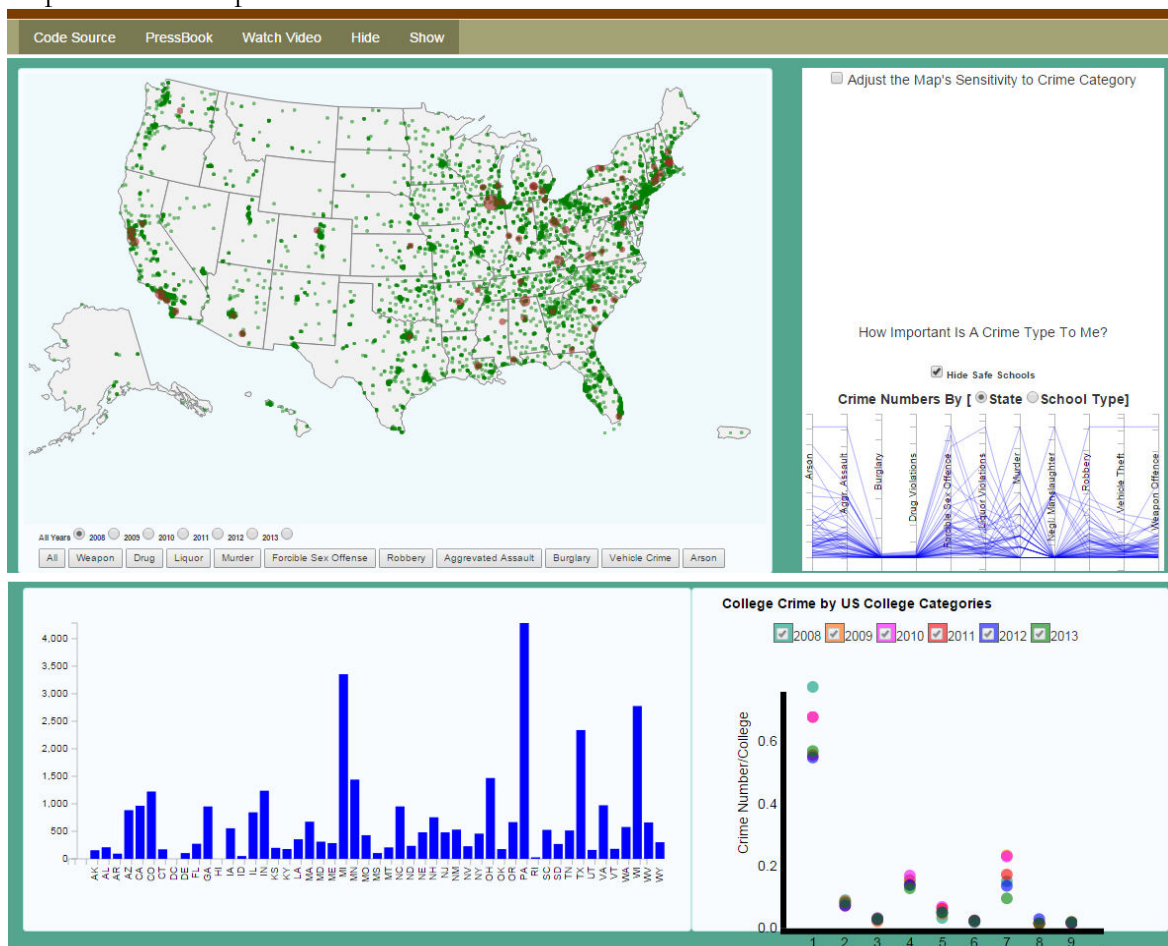


Fig 9

VI. IMEPELETION OF THE DESIGN

Site Structure and Responsive Design

(April 8-9, April 11 Linghong)

The site main structure is a map that function as visually displaying college on campus security data. Alongside the map, there are a few boxes, which act as either a user interactive area to control the map view, or a place to show the college security patterns or trends. The boxes size, color or position will be changed according to the content size and visualization purpose.

The whole site (including the map) is made responsive to screen size change, and can be viewed in desktop, i-pad and smart phone with a layout suitable to each screen size. Especially, when viewing the site in a smart phone, the menu button will be collapsed. Clicking the button, the menu will show up. Bootstrap are used to organize the site structure, look and the responsive design.

Interactive Control Panel and Weighting Method

(4.22, 5.1-5.5, Suhas)

All visualizations will get the data from data/crimeData.json. This file will have raw crime data in the following structure:

A small json sample is available in data/sample.json.

```
{ "type": "crimeData", "schools": [
{
  "school": {
    "schoolId": "100654001",
    "name": "Alabama A & M University",
    "branch": "Main Campus",
    "address": "4107 Meridian St",
    "city": "Normal",
    "state": "AL",
    "zip": "35762",
    "latitude": "34.733964",
    "longitude": "-86.522913",
    "sectorCd": "1",
    "yearData": [
      { "yearOfData": "2008",
        "murderCount": "0",
        "negligentManSlaughter": "0",
        "forcibleSexOffense": "1",
        "nonForcibleSexOffense": "0",
        "robbery": "4",
        "aggravatedAssault": "26",
        "burglary": "122",
        "vehicleTheft": "3",
        "arson": "1",
        "weaponOffense": "4",
        "drugViolations": "9",
        "liquorViolations": "1" },
      { "yearOfData": "2009" /* rest of the fields will repeat here* / },
      { "yearOfData": "2010" /* rest of the fields will repeat here* / },
      { "yearOfData": "2011" /* rest of the fields will repeat here* / },
      { "yearOfData": "2012" /* rest of the fields will repeat here* / },
      { "yearOfData": "2013" /* rest of the fields will repeat here* / }
    ]
  }
}
```

```

    }
  }
  {"school":
  {"schoolId":"100663001"/*school data will repeat here*/}}}}
}

```

To process this raw data, a new js file `crimedataAnalyzer.js` is created. The constructor of this file received the raw json data and it is expected to process all “views” of the data for the individual parts of the visualizations.

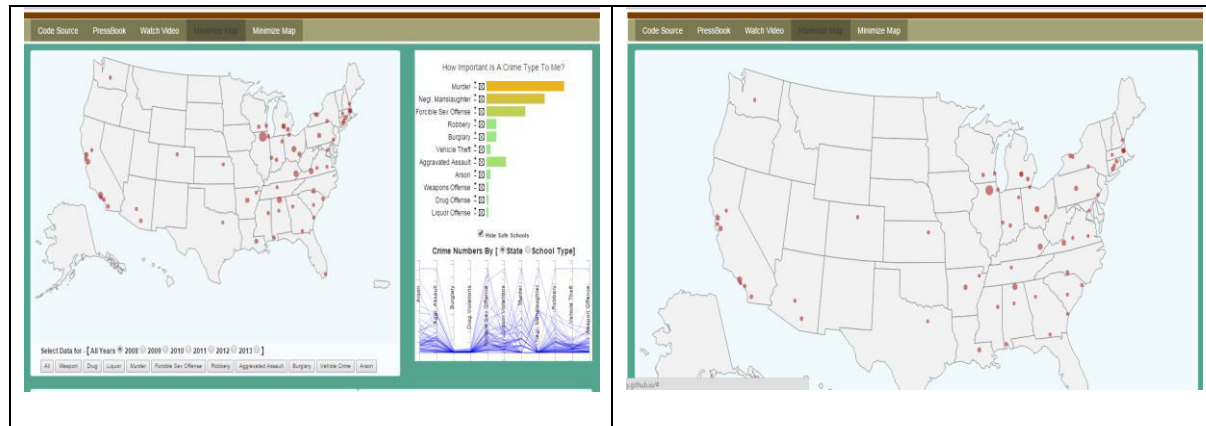
For example, the map requires the data aggregated by year and data for all years. In the `init` method, it populates `yearCrimeData` and `allTimeCrimeData`.

For the states data, following structure will be used:

```

{
  "type": "stateCrimeData",
  "states": [
    {
      "state":"MD",
      "yearCounts":
      {
        "years":{
          "*":[
            {
              "murder":10,
              "vehicleCrime":50
            }
          ],
          "2006":[
            {
              "murder":10,
              "vehicleCrime":50
            }
          ],
          "2007":[
            {
              "murder":10,
              "vehicleCrime":50
            }
          ]
        }
      }
    },
    {
      "state":"DE",
      "yearCounts":
      {
        "years":{
          "*":[
            {
              "murder":10,
              "vehicleCrime":50
            }
          ],
          "2006":[
            {

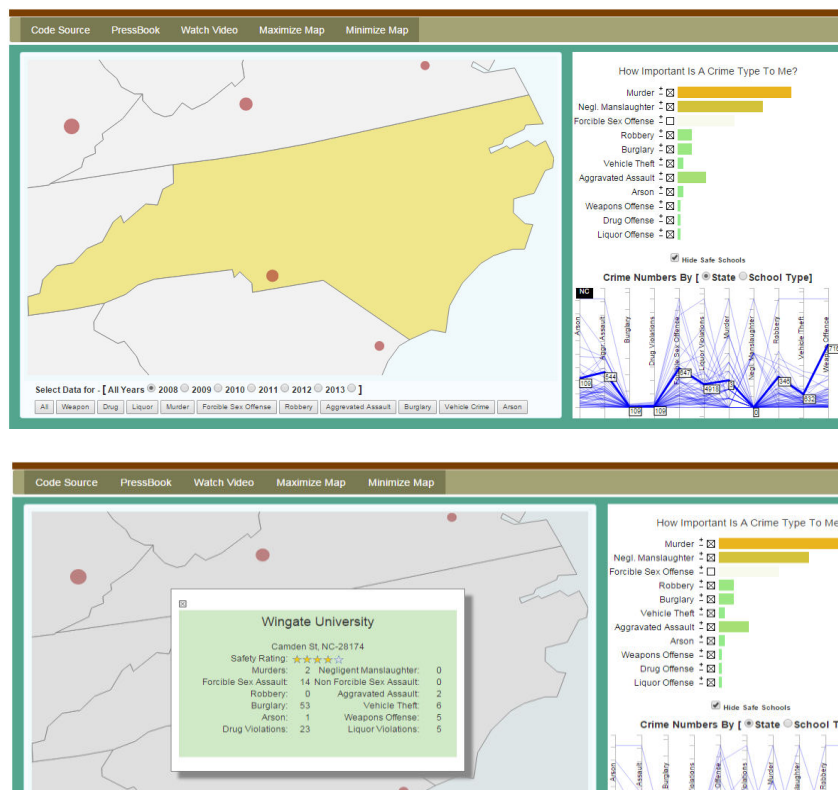
```

Map View

(April-May 5th Suhas)

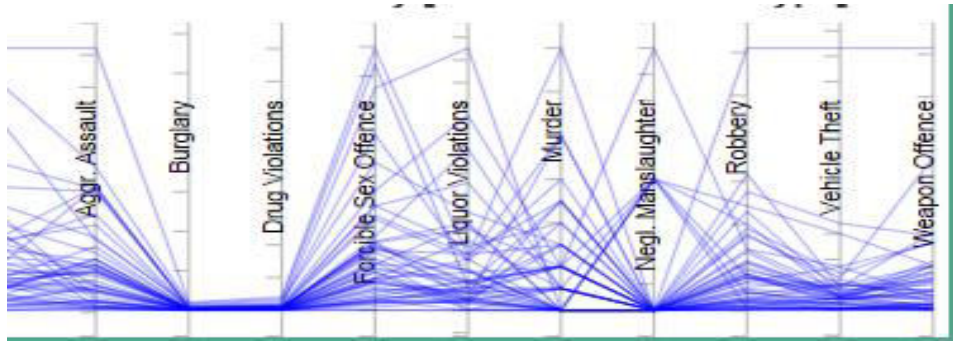
The map has two views, a full map view, or zoom to just one state. Hovering to that cycle, a college information panel will show up.



Parallel Coordinates

(May1- 5th, 2015, Suhas)

A Parallel Coordinates is used to show the crime volume in each state or each school type. The Parallel Coordinates in the state level also serves as a control panel. Hovering to a state in the parallel coordinate will make the map to zoom to that state.



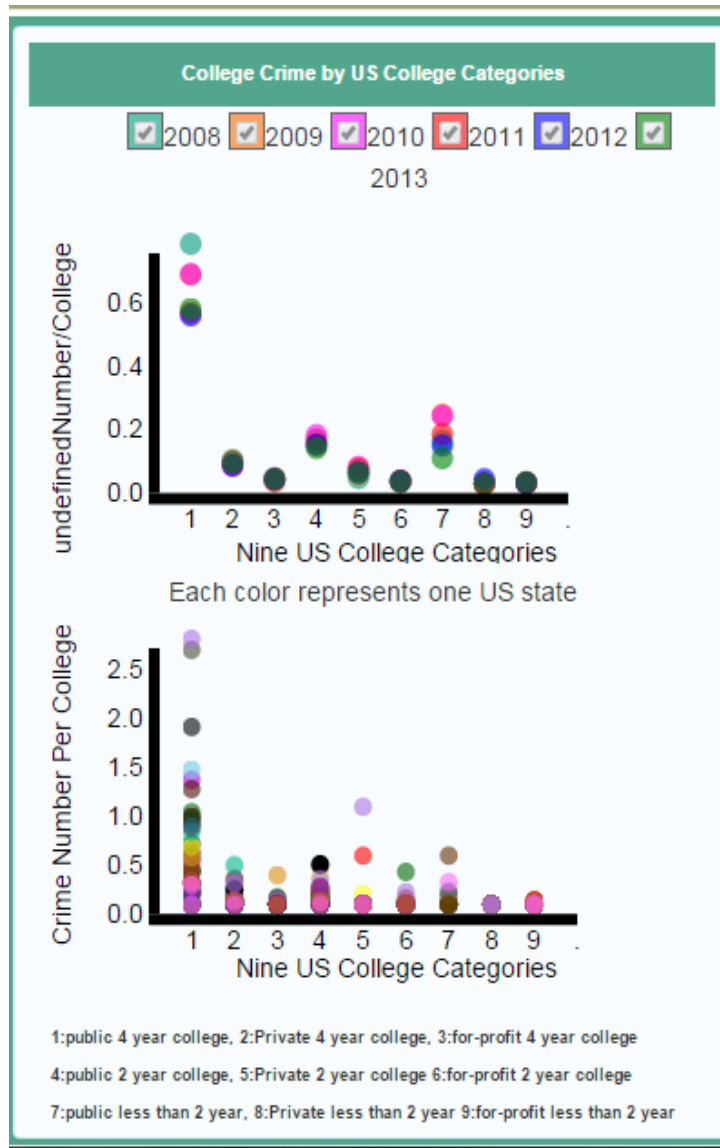
Multi-Series Scatter Plots

(May 5, 2015, Linghong)

To show users that the on campus school crime is college type related, two multi-series scatter plots were made. The x-axis indicates nine college categories, while the y-axis indicates the crime ratio(cases/per college). From the fig, we can see public four year colleges has much higher crime ratio, and these patterns are consistent across different years. While for other college types those crime number are almost same each year, meaning it is a predictable reaction. Finding a pattern is much more meaningful than showing individual data. You will quickly forget the number of a particular crime, but you can easily remember the fact that public four year colleges have much higher crime rate than other colleges types.

Form the state level graph(the lower panel), we see the same distinct college type related crime pattern, which is reassure us that the trend is real. However, in the state level, higher variation of the crime ratio are seen among different states. In the public 4 year college categories, we have large range of variation across different states, but in other college categories, we see some ridiculous high crime rate outlays.

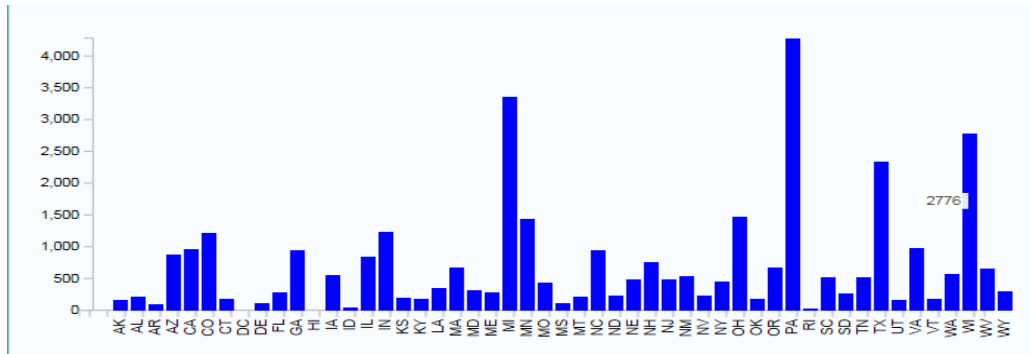
Clicking crime category buttons and the time radio buttons beneath the map will trigger the changes in both scatter plots. The colored check boxes is used to select various years. However, the final map doesn't include the scatter plot charts of the lower charts to avoid bringing confuse to users.



Barcharts

(May 5, 2015, Peter)

A barcharts is used to show crime volume in 50 states. The bar charts show clear large variation across different states.



VII. EVALUATION

(May 5th, Linghong)

- What did you learn about the data by using your visualizations?

From the work we have done, we unexpectedly found the public colleges have much higher crime number. However, we also see in majority of crime categories, the crime number for public 4 year colleges goes down year by year. This view is not shown in the site due to time limit. Please view it in the data analysis section.

From the site, we can also view a college, a state or any college type's crime level. We also get general picture how are the crime pattern distributed across US.

- How did you answer your questions?

For the questions we asked in the beginning, we are happy about what we have achieved. However, due to the time limitation, several of the questions leave unanswered. For example, the college application and college ranking data file has only less than two hundreds of colleges, and they are also all stays in the very top of the college ranking. We think it is probably unlikely we will see a pattern with crime correlation by only using schools in the top categories. A data set spreading across different college quality levels would potentially be more useful.

Another issue is after analyzing our data we realized that the student number, the gender ratio and the total students from the original data set were only provided in a university basis, not the campus basis. Since about 10% of colleges didn't submit the crime data. This made analyzing of the data using any students number impossible.

- How well does your visualization work, and how could you further improve it?

We have achieved our initial goals. Now we have a map view with two very different purposes of the usage. Those different views are nicely connected and interact well with the three Visualization boxes. We also have some analysis for crime patterns. Especially, we found a distinctive college-type-related crime patterns, which is

unexpected and exciting. Also thinking about the large crime variation across various locations disclosed through our map, the 50 states bar chart, and the scatter plot we didn't show, all of these indicate the same point.

Since 4 year public college has large variation across different states, we did a quick analysis of using the tool of parallel coordinates. We compared states level 4 year public colleges' 2013 crime data with the median household income, and the state unemployment ratio. However, only those drug and weapon levels seemed slightly correlated to those crime data, but for other crime categories, we didn't find very obvious relation. If we have more time, the cause of the high crime ratio in 4 year public is certainly something we want to dig further, and show in our vis..